DOCUMENT RESUME

ED 159 201 95 TM 007 469

AUTHOR	Shaycoft, Marion F.
TITLE	Project TALENT: A Short History of a L--o--n--g Project [and] Technical Appendices.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	AIR-FR-55400-9/77
PUB DATE	Sep 77
GRANT	NIE-G-74-0003
NOTE	460p.; Parts are marginally legible due to print quality
EDRS PRICE	MF-$0.83 HC-$24.77 Plus Postage.
DESCRIPTORS	Academic Achievement; Aptitude; Biographical Inventories; Career Choice; Data Analysis; Data Bases; Educational Assessment; *Followup Studies; *High School Students; Longitudinal Studies; *National Surveys; *Program Descriptions; Research Design; *Research Methodology; Research Problems; Senior High Schools; *Student Characteristics; Testing Programs; Test Selection
IDENTIFIERS	*Project TALENT

ABSTRACT

In 1959, the American Institutes for Research (AIR) and the University of Pittsburgh selected a stratified random probability sample of all secondary schools (public, parochial, and private) in the United States, tested all their students in grades 9-12 with a two-day battery of aptitude tests, information tests, tests of ability and achievement, and inventories probing the examinees' background, interests, and personality traits. Following the initial testing, examinees were periodically surveyed through a series of mailed questionnaires; this effort has yet been completed. The study was labeled Project TALENT, and was designed to provide: (1) an inventory of human resources; (2) a set of standards for educational and psychological measurement; (3) a counseling guide indicating the patterns of aptitude which are predictive of success in various careers; (4) a better understanding of how young people choose their life work; and (5) a better understanding of the educational experiences which prepare students for their life work. The project is traced from its inception in 1956 to its current status, including future follow-up plans. This informal history includes information concerning project staff, reports, test batteries, retest studies, data tapes, computer programs, statistical and management problems, and follow-up questionnaires. Statistical analyses are discussed in the technical appendices. (Author/JAC)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ENT OF HEALTH.
N & WELFARE
INSTITUTE OF
CATION

HAS BEEN REPRO-
AS RECEIVED FROM
RGANIZATION ORIGIN-
OF VIEW OR OPINIONS
NECESSARILY REPRE-
ATIONAL INSTITUTE OF
TION OR POLICY

AIR-FR-55400-9/77

TM

Project TALENT: A Short History of a L--o--n--g Project

MARION F. SHAYCOFT

MERICAN INSTITUTES FOR RESEARCH
ost Office Box 1113 / Palo Alto, California 94302

AIR-FR-55400-9/77

Project TALENT: A Short History of a L--o--n--g Project

MARION F. SHAYCOFT

This work was done under Grant No. NIE-G-74-0003 from the National Institute of Education, Department of Health, Education, and Welfare. The content does not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

AMERICAN INSTITUTES FOR RESEARCH
PALO ALTO, CALIFORNIA

1977

PREFACE

A few words about the nature and purpose of this book

This is an "internal" history of Project TALENT. In other words, it is not about the nearly half a million people "out there" to whom we administered two days worth of tests when they were high school students in 1960 and who are now in their early 30's. No, it is not about them and what has happened to them in the years since 1960; there are already a great many massive research reports telling about that. What the "internal history" focuses on primarily is what happened at places like:

- 410 Amberson Avenue, Pittsburgh
- 6135 Kansas Avenue, Washington
- 1808 Adams Mill Road, Washington
- 200 South Craig Street, Pittsburgh
- 135 North Bellefield Avenue, Pittsburgh
- 1791 Arastradero Road, Palo Alto

(where Project TALENT has had its headquarters at various times), and at places in Washington like:

- 400 Maryland Avenue S.W.
- 1200 19th Street N.W.

(where the Office of Education and the National Institute of Education respectively have their headquarters).

In the 20 years that have elapsed since the project was just a gleam in John Flanagan's eye, a lot has happened and we have learned a lot. If it is true that one learns by experience, then one should learn a lot in a long-term longitudinal study--just for the very reason that the study does last so long and has so many facets. One of the ideas behind this History is that it would be nice if future researchers--particularly researchers doing large-scale longitudinal studies--could learn from our experience.

In a period as long as 20 years, it is inevitable that we would have made some mistakes. But we must have done many things right, too; otherwise the project would have collapsed and died long ago. In this History then, we are trying to take a long retrospective look, so that we can tell our readers about our errors and our successes, in the hopes that future researchers will avoid the former and may see fit to emulate the latter.

A literary genre known facetiously as the "HIBK murder mystery" was popular at one time. This was a murder story written in the first person singular, in which the narrator, usually an amateur detective, frequently started statements with "Had I but known [HIBK] then what I know now. . . ." I mention this because the present History, though not a murder mystery, also partakes, to some extent, of the HIBK feature, at least by implication, every time it contains a discussion of errors made, or things done in less than the optimal way.

Inevitably, in discussing what was done right and what could have been done better, a lot of personal opinion must enter the picture. That is an inherent feature of a History such as this. Whenever opinions are expressed in this volume, they are my opinions unless explicitly attributed to someone else. This is, as someone who read the manuscript commented, a "personal document." Given the nature of the assignment and the function of the History, I think that that is inevitable. And some of these opinions expressed in this "personal document" may be wrong; almost certainly some of them are wrong, since nobody is infallible. But since I wouldn't want my errors attributed to someone else, I repeat: the opinions I have expressed here are mine. Other people may or may not share them. Some readers may be bothered by the fact that some of the opinions expressed are not exactly dispassionate, but I know of no way that it could be avoided.

And if there is an apparent undue emphasis on my own role in the project, I apologize for that. I certainly have tried not to slight anyone else's role, or to give the impression that my role has been more important than it really was. Thus if there seems to be an undue emphasis on my role it is because that is what I know best; and if there is an undue emphasis on my own viewpoint, it is because that is what I understand best.

Nevertheless it is at least a little embarrassing that the Technical Appendices turned out to consist almost exclusively of my own writings. This should not suggest that the author of this History is the only staff member to have written papers or presented papers at professional meetings. Quite the contrary! As indicated in Chapter 21, over the years more than 100 papers have been presented by staff members at professional meetings and a very large number of staff members have been among their authors. But the author of this History has specialized in methodological problems to a somewhat greater degree than most of the other staff members and therefore has had more occasion to write papers with methodological implications. The Technical Appendices are intended largely as a technical guide to various rather special aspects of research with TALENT data, sometimes requiring special methodological twists. For a long time I was virtually the only one on the project staff working on problems of that sort rather than almost exclusively on substantive problems (as were most of the other members of the research staff). And that is how it happens that the author of this History is also the sole author of most of the Appendices.

The period covered

A History needs a cut-off point. This History covers the events occurring from the inception of TALENT in 1957 through the final processing of the 11-year follow-up data in early 1977. But readers (and authors) may find a cut-off date frustrating when they know that things must have happened after that date. Therefore a "Postscript" has been appended after the last chapter, to bring things up to date.

How to find your way around in this book

I suspect that this volume will tell some readers more about Project

TALENT than they are interested in learning. Since I have no wish to burden the reader with unwanted information, there is a fairly detailed Table of Contents. This, together with the usual "List of Tables" and "List of Figures" will, it is hoped, let the reader with specific information needs find what he is particularly interested in at a particular time.

There is also a glossary of acronyms and abbreviations which the reader may find helpful if he stumbles on a mysterious jumble of letters, such as AFPTRC or NCERD or T.E.A.R.S. The glossary is the last thing in this volume.

Most of the appendices are bound together in a separate volume.

A few words on nomenclature

In writing this book the author encountered a couple of slight problems of nomenclature. The first one concerned how she should refer to present and former colleagues on Project TALENT. Five options clamored for consideration: (1) full name as used officially in a professional context (e.g., John C. Flanagan, A. Carp, William V. Clemans, Dorothy S. Edwards); (2) first and last names as used conversationally to refer to the person named (e.g., John Flanagan, Al Carp, Bill Clemans, Jo Edwards); (3) last name with honorific title attached (e.g., Dr. Flanagan, Dr. Carp, Dr. Clemans, Dr. Edwards); (4) last name only (e.g., Flanagan, Carp, Clemans, Edwards); (5) first name only (e.g., John, Al, Bill, Jo). Options 2 and 5 have the advantage of informality--which fits in with the fact that this is a somewhat informal history. Option 5 corresponds to the prevalent mode of address but it can lead to ambiguity except where context helps. (Does "Bill" mean Bill Gorham? Bill Cooley? Bill Clemans? And does "Jo" really mean "Dorothy"?) Option 3 is somewhat more formal and certainly less subject to ambiguity. Option 4 corresponds to the standard mode of reference used in professional articles. If the problem posed above were a five-option multiple-choice item it would be a rotten one; it doesn't clearly have one and only one correct answer--nor even one and only one best answer.

The solution chosen was to use whichever style of reference seemed most appropriate or most convenient in the context in which it appeared. This means that almost every one of the five options was used at one point or another. The author takes what comfort she can from Ralph Waldo Emerson's well-known comment on consistency.*

That brings us to the second nomenclature problem. How should the author refer to herself? As "the author"? Or by name? Or in the first person plural? Or first person singular? Again I invoke Emerson's comment on consistency!

Marion F. Shaycoft

September 1977

---

* "A foolish consistency is the hobgoblin of little minds."

ACKNOWLEDGMENTS

There are two people whose contributions to this effort were such that without them this book couldn't have been--or wouldn't have been--written. Chronologically Dr. Andrew C. Porter is the first, since it was he who really stimulated the idea, when he was directing the Basic Skills Group at the National Institute of Education. (He is now on the faculty of Michigan State University.) As a result ot a proposal that the author of this History submitted to the Basic Skills Group, which bore on the topic of the desirability of disseminating what we had learned in Project TALENT about running a large-scale longitudinal study and processing and analyzing the data from such a study, Andy had the idea that all of that could be incorporated in a history of Project TALENT as viewed from the inside. That idea of Andy's led, eventually, to a grant, and that grant resulted in this History. Thus it is clear that without Andy's interest and action this History would not have been written.

After the grant was awarded and I started working on the History, Andy maintained close contact and a continuing interest. He reviewed the outline for the History and made suggestions on it; I discussed new developments with him; and his interest in it has continued to this day.

In an early conference, Andy urged me to write in an informal style rather than in the style of a typical research report. I don't know whether this History has turned out in line with his hopes, but I am absolutely certain that it doesn't sound like a typical research report!

The other person whose contribution to this History has been of critical importance is John Flanagan. Certainly without the information he provided about events that occurred before the author of this History joined the Project TALENT staff and about other events that were going on soon after that, most of Chapter 2 couldn't have been written.

Furthermore througnout the writing process he has been a sounding board. I have checked recollections with him, tried out ideas on him, and given him draft copy to review. That review has proven most helpful--as were the

numerous conferences at which he functioned as an informal in-house consultant. Thus it is clear that without John's very substantial contribution this History probably couldn't have been written; or if it had been, it would have looked quite different. It would have been a lot shorter, being truncated at the front end. And the errors that John caught in reviewing the draft wouldn't have been caught by anybody.

John Dailey, who was the Project Director from 1958 to 1964 is another person who has been extremely helpful. I have conferred with him frequently, both in person and by phone, and he too has helped to fill the gaps in my information about some of the things that happened in the early days of the Project. He was kind enough to review the first half of the manuscript at my request.

Jo Edwards of the AIR staff in the Washington Office, who was also on the TALENT staff in the early days of the project, was also kind enough to answer lots of questions I threw at her about her recollections, and to review the first half of the manuscript. She provided many helpful suggestions.

One other person whom I was able to consult who was associated with the project in its early days, and later too, was Alice Scates, of the Office of Education. She was the project's first monitor, and like all the other people I have mentioned, she was very helpful.

It must be clear from the foregoing that while this History is based to a substantial degree on documented evidence, there are many parts that are dependent on memory--mine and others'. Thus it wouldn't be at all unlikely that there are some errors of fact--but I imagine there are very few since we have tried to check everything checkable, and since, reassuringly, my consultants' memories agree with each other in general.

Jack Schwille, who until recently was at the National Institute of Education, continued Andy Porter's interest in the developing "History" after Andy left NIE.

I cannot end this Acknowledgments section without saying thank you to the two staff members--Rachel Holmen and Nancy Hull--who have done the most to convert this history from a completely handwritten partly illegible scrawl into a finished report. Rachel has been associated with this effort for about a year, and in addition to her typing skills she has provided very competent assistance in other aspects, by functioning also as a research assistant and administrative assistant. Nancy Hull entered the picture later, but she too has provided help and facilitated the effort in many ways besides typing.

My thanks to all the people mentioned above.

Of course I take responsibility for any errors of fact in this manuscript, and as I hope the Preface has made clear, if there are any errors of opinion those errors are mine alone.

Marion F. Shaycoft

September 1977

TABLE OF CONTENTS

Page

Preface ii
Acknowledgments vii
Table of contents x
List of tables xviii
List of figures xix

Chapter

1. Prologue 1- 1
    A. What Project TALENT is 1- 1
    B. Frame of reference 1- 2
    C. Current status of the Project 1- 3
    D. In the public eye! 1- 3

2. The Prenatal Stage: (Project TALENT Before It Was Project TALENT) . . . 2- 1
    A. Roots of the idea 2- 1
    B. The start of the search for funding 2- 4
    C. The planning and the plans 2-10
    D. Finding funding 2-12
    E. Bonfires and brouhahas 2-13
        1. The Baltimore brouhaha 2-14
        2. The Texas bonfire 2-15

3. Issues and Decisions 3- 1
    A. The sample 3- 1
        1. The Sampling Advisory Panel 3- 2
        2. After the Sampling Advisory Panel meeting . . 3- 7
        3. The 16-year-olds and the 15-year-olds . . . 3- 9
        4. The 10 subsamples: an interim feature . . . 3-10
    B. The design 3-11
    C. To build or buy: The TALENT Battery 3-12

4. Naming the Baby 4- 1

5. Building the Battery 5- 1
    A. Schedule 5- 1
    B. "State-of-the-art" 5- 1
    C. Content and organization of the battery 5- 2
    D. The battery-building crew 5- 5
        1. The test builders 5- 5
        2. Questionnaire for students 5- 8
        3. School questionnaire 5- 8
    E. Tryout and item analysis 5- 9
    F. The end of the process 5- 9
    G. What we ended up with 5-13

Chapter Page

6. Pre-"Criterion-Referenced" Criterion-Referenced Tests (The TALENT Domain Tests) 6- 1
    A. The concepts 6- 1
    B. The tests 6- 2

7. Execution of the Plan 7- 1
    A. Selecting the sample 7- 1
    B. Getting the cooperation of the selected schools . 7- 3
    C. Administering the tests 7-12
    D. E. F. Lindquist and Measurement Research Center . 7-19
    E. Implementation 7-25
        1. Before MRC (The "TALENT Laundry") 7-25
        2. At MRC 7-27
        3. After MRC 7-27

8. With a Cast of Thousands! 8- 1
    A. The staff 8- 1
        1. The Principal Investigator 8- 5
        2. Project Directors 8- 5
        3. Associate Directors 8-11
        4. Data Bank Directors 8-12
        5. Other recent senior staff members 8-13
        6. Personnel in charge of computer operations . . 8-14
        7. Computer trouble-shooter task forces 8-16
        8. Other research staff members 8-18
            a. Washington phase 8-18
            b. Second Pittsburgh phase 8-20
            c. Palo Alto phase 8-21
            d. Washington, Pittsburgh, and Palo Alto . . 8-22
    B. Early advisory panels 8-23
    C. Regional Coordinators 8-23
    D. OE and NIE personnel 8-23
    E. TALENT post-doctoral fellows 8-26

9. The Geography of the Project 9- 1

Chapter Page

10. Early Reports 10- 1
    A. Basic reports 10- 1
    B. Equating studies 10- 2
    C. Subsequent studies applying the results of the equating 10- 2

11. Returning to the Scene 11- 1
    A. Designing the retest study 11- 1
        1. Who 11- 1
        2. Where 11- 1
        3. What 11- 2
        4. When 11- 2
        5. How much 11- 2
        6. How 11- 3
    B. Processing the answer sheets 11- 3
    C. Data analyses and final report 11- 3
    D. Test reliability data 11- 4

12. The 4% Sample 12- 1

13. Tapes: Trauma, Troubles, Travail, Trials, Tribulations, and T.E.A.R.S. 13- 1
    A. TALENT's tapes: Ancient history 13- 1
    B. 1. Iowa City (1960) 13- 1
        2. Coping with geography 13- 1
        3. Zeros and blanks 13- 2
        4. Building a tape master file 13- 3
        5. Design of the tape master file 13- 6
        6. Other instances of data loss 13- 6
    B. Interim Alphabetic Directory 13- 8
    C. The concept of follow-back studies 13-10
    D. Tape editing and reformatting system (T.E.A.R.S.) 13-11
        1. The NIE panel 13-11
        2. The proposal 13-11
        3. September 1975 - April 1976 13-12
        4. April 1976 - May 1977 13-19

14. Computers and Computer Programs 14- 1
    A. Problems with programs 14- 1
    B. Keeping up with technology 14- 2

Chapter Page

15. Struggling with Weights (Not a Chapter About Weight-Lifting) 15- 1
    A. Sets of weights that have been developed . . . 15- 1
        1. Weights for 1960 data (where no follow-up data are being used) 15- 1
        2. Weights for follow-up data 15- 3
    B. Whether and how to weight 15- 5
    C. Self-weighting: A solution to the trade-off dilemma 15- 6

16. Other Statistical Problems Encountered and Solved . . 16- 1
    A. Sampling 16- 1
        1. How to handle schools that are in the project on a specially invited or volunteer basis . . 16- 1
        2. The "feeder junior high school" problem . . 16- 2
        3. Coping with complexly structured samples in computing standard errors 16- 2
        4. Is the Special Sample of nonrespondents necessary? (An empirical validation of follow-up Weight C) 16- 2
        5. Can the selection of the Special Sample be improved? 16- 3
    B. Reducing sampling error--both random and systematic 16- 3
    C. Reliability coefficients and related statistics . 16- 4
        1. Reliability coefficients 16- 4
        2. Coefficients of internal precision 16- 5
    D. Inferring score distributions for an unselected age group 16- 6
    E. Problems in methodology related to career guidance 16- 6

17. Management Problems 17- 1
    A. Record-keeping problems 17- 1
        1. Identification numbers and other codes . . . 17- 1
        2. Some comments about redundancy 17- 1
        3. Confidentiality 17- 3
        4. Retrievability 17- 3
    B. Securing and maintaining cooperation 17- 3
    C. Getting funding (the cliff-hanger serial) . . . 17- 5

Chapter | Page

17. Management Problems (continued)

- D. Problems 17- 9
  - 1. Turnover 17-11
  - 2. Personnel selection 17-11
- E. Management problems related to computer and computer-related operations 17-12
  - 1. Management of keypunching operations . . . 17-12
  - 2. Coordination between the research staff and the computer staff 17-13
  - 3. Backup tapes 17-14
- F. Pitt leaves TALENT before TALENT leaves Pittsburgh leaves Pittsburgh! 17-14

18. The follow-ups 18- 1

- A. Mailing of the questionnaires 18- 1
- B. Processing the returned questionnaires 18- 2
  - 1. Keypunching vs. optical scanning 18- 2
  - 2. Coding 18- 4
  - 3. Name and address changes 18- 4
- C. Mail response rates 18- 5
- D. Special samples of nonrespondents 18- 6
- E. Efforts to maintain and improve response rates . 18-10
  - 1. For mailed questionnaire 18-10
  - 2. For special sample 18-11
- F. Content of special-sample questionnaire 18-13
- G. Ten becomes eleven; twenty equals seventeen? . . 18-14

19. TALENT Post-Doctoral Fellowships 19- 1

20. Later Studies 20- 1

- A. The one-year follow-up studies 20- 1
- B. The MAP studies 20- 1
- C. The 1963 retest study 20- 1
- D. The five-year follow-up studies 20- 3
- E. Ted Cureton's two TALENT studies 20- 4
- F. The Career Data Book and the supplement to it . . 20- 4
- G. The 11-year follow-up studies 20- 5

Chapter Page

21. Some Other Uses of TALENT Data 21- 1

A. Equating studies 21- 1

1. For the Air Force 21- 1
2. For the Navy 21- 1
3. Knoxville equating 21- 2
4. Inter-American Tests 21- 2
5. PCG equating 21- 2

B. Knoxville follow-up studies 21- 2

C. Other studies of marriage, the family, and parenthood 21- 2

D. Changes in norms 21- 3

E. Quality-of-life studies 21- 3

F. Follow-back studies 21- 4

G. A guidance study for NSF 21- 4

H. Sex differences in achievement 21- 4

I. Volunteer military forces 21- 5

J. Project PLAN 21- 5

K. Data Bank studies 21- 5

L. Miscellany 21- 5

M. Undone studies 21- 6

22. The Data Bank 22- 1

A. History of the Data Bank 22- 1

B. Extent of use 22- 2

C. Charging for Data Bank services 22- 4

D. Policy on use of TALENT Battery 22- 4

E. Policy on use of Data Bank services 22- 5

1. Maintaining confidentiality 22- 6
2. Raw data tapes versus data analysis printouts 22- 6

F. Handbooks, brochures, etc. 22- 8

23. Bridging the Gap: The PCG (Planning Career Goals) Battery 23- 1

A. Resolution of the dilemma 23- 2

B. Planning the battery 23- 3

D. Developing the PCG Battery and equating it to TALENT 23- 5

Postscript . . . . p- 1

Page

References . . . . r- 1

Appendix A. The TALENT Battery . . . . A- 1

  Appendix A1. Composition of the Project TALENT Battery . . . . Al- 1

  Appendix A2. Thumbnail descriptions of the Tests . A2- 1

Appendix B. Choosing Which 1960 Variables to Use . . . B- 1

Glossary of Acronyms and Abbreviations . . . . g- 1

TECHNICAL APPENDICES
(separately bound)

Page

Appendix C. The Coefficient of Internal Precision: A New Psychometric Statistic for Heritability Research . . . . C- 1

Appendix D1. Must All Members of a Probability Sample Be Randomly Selected? D1- 1

Appendix D2. The "Closed School-Cluster" Method of Selecting a Probability Sample D2- 1

Appendix E1. When is Bias Better? E1- 1

Appendix E2. A Stratified Procedure to Compensate for Nonrespondents E2- 1

Appendix E3. Minimizing Total Error: A Trade-Off Problem E3- 1

Appendix F. Statistical Procedure for Inferring Score Distribution of Unselected Group of 18-Year-Olds from Score Distributions of Unselected Group of 15-Year-Olds and Twelfth-Grade Students . . . . F- 1

Appendix G1. A New Multivariate Index for Use in Educational Planning G1- 1

Propinquity Analysis: A New Guidance Tool . G1-20

Appendix G2. The Project TALENT Approach to the Career-Grouping Problem: Cutting the Gordian Knot G2-1

LIST OF TABLES

Table No.		Page
1-1	Project TALENT original follow-up schedule . . .	1- 4
3-1	Sampling ratios and senior class size strata . .	3- 9
15-1	Taxonomy and etiology of statistical error . . .	15- 2
17-1	TALENT funding for first 20 years (1957-76) . . .	17-10
18-1	Mailing schedule and procedure for 11-year follow-up of grade 10 cohort, and TALENT News . .	18-12
23-1	List of the occupations included in each of the the 12 career groups	23- 4
23-2	PCG and corresponding TALENT instruments	23- 6

LIST OF FIGURES

Figure No.		Page
1-1	Remarks by "Congressional Record" of January 20, 1960	1- 6
1-2	Letter from Vice President Lyndon B. Johnson to John T. Dailey, March 19, 1962	1- 9
2-1	Advisory Panels	2- [illegible]
2-2	Research Advisory Committee of the Cooperative Research Program (1958-59)	2- 9
2-3	Advisory Council	2-11
3-1	Sampling Advisory Panel	3- 3
5-1a	Record Form Z (side 1)	5-10
5-1b	Record Form Z (side 2)	5-11
5-2	Extract from Local Coordinators' Manual	5-12
7-1	List of schools in sample as of 22 June 1960	7-13
7-2	The national sample of schools [Map]	7-17
8-1	Project professional, technical, and supervisory personnel: 1957-76	8- 2
8-2	The changing cast: 1957-76	8- 4
8-3	Regional Coordinators in 1960	8-24
9-1	The geography of Project TALENT (Who and what were where when)	9- 3
13-1	Priority chart for each additional step	13-13
13-2	Description of Part A of tape documentation	13-15
13-3a	A Part B form intended primarily for test scores	13-17
13-3b	A Part B form intended for other kinds of information including questionnaire items	13-18
13-4	Project TALENT data file format	13-21

Chapter 1. PROLOGUE

If you were born between 1942 and 1945 and went to high school in the United States, there is roughly one chance in 25 that you were a participant in Project TALENT, maybe you were a participant even though you were born a little earlier or a little later--say in 1941 or 1946.

But even if you weren't a participant it is quite likely you know someone who was (although he or she may never have had occasion to mention it to you). And if you don't, the chances are still pretty good that you know someone who is or was on the staff of the project or who was or is associated with one of its sponsors. The fact that you are reading this book suggests that you are a member of the educational research community. Over the years there have been approximately _____ professional staff members, and in the course of the nearly 20 years since the beginnings of the project those _____ people have scattered widely. There have been innumerable college undergraduates and graduate students in the Palo Alto, Pittsburgh, and Washington, D.C. areas who have worked part-time as coding clerks or other junior staff members, many of whom eventually sought careers in educational research. They too have scattered widely. Therefore, to reiterate, perhaps you know someone who is or was on the Project TALENT staff. Chapter 8 (which has the somewhat hyperbolic title "With a Cast of Thousands!") tells a bit about who some of those people are.

A. What Project TALENT is

Project TALENT is a very-large-scale jointly cross-sectional and longitudinal educational research study that some people may think has been going on forever. It hasn't. But it was going on as far back as 1957, if one counts the planning phases that were carried out before the project really began to take shape in anything resembling its present form. And it had its start even further back than that, if one traces it back to when the idea was born. Actually in Chapter 2 we trace the idea even further back than that, to its "prenatal stage."

To condense the description of Project TALENT into a single sentence: In 1959, primarily under the sponsorship of the U.S. Office of Education, the American Institutes for Research (AIR) and the University of Pittsburgh jointly undertook a massive project, which involved scientifically selecting a stratified random probability sample of all secondary schools (public, parochial, and private) in the United States, testing all their students in grades 9-12 (nearly half a million students) with a two-day battery of aptitude tests, information tests, other tests of ability and achievement, and inventories probing the examinees' background, interests, and personality traits, and then following up the examinees by mailed questionnaire periodically over a very long period of time (which hasn't ended yet). There is no point in expanding that very condensed description of Project TALENT at this point, because that is what most of the rest of this book is about.

B. Frame of reference

A frame of reference for the project is provided by its original purposes and its original schedule. The purposes of the project, as originally stated, were fivefold:

a. To provide an inventory of human resources.

b. To provide a set of standards for educational and psychological measurement.

c. To provide a comprehensive counseling guide indicating the patterns of aptitude and ability which are predictive of success in various careers.

d. To result in a better understanding of how young people choose their life work.

e. To result in a better understanding of the educational experiences which prepare students for their life work.

The original plan was to follow up each of the four grade cohorts four times--one year, five years, ten years, and twenty years after they were scheduled to graduate from high school. Table 1-1 shows the schedule.

C. Current status of the Project

The one-year and five-year follow-ups were completed on schedule. As things turned out, however, the 10-year follow-up did not occur. Because of a one-year delay in funding, we conducted an 11-year follow-up instead.

The 10-year follow-up had been scheduled to start in 1970. Instead, the 12th grade cohort was followed up in 1971. It was about this time that plans for the creation of the National Institute of Education (NIE) were being worked out. Project TALENT was among the activities scheduled to be transferred to NIE whenever it came officially into existence. It is probable that uncertainties associated with the imminent creation of NIE and the planned transfer of some of the functions of the Office of Education (OE) to the new Institute had something to do with the delay in funding.

Table 1-2 shows the schedule on which follow-ups have been completed thus far.

It perhaps should be mentioned at this point, although to do so involves breaching the definition of a "history" by talking about the future, that the possibility of another major change in the schedule has been broached recently (by NIE personnel). It has been suggested that the 20-year follow-up be changed to a 17-year follow-up.

D. In the public eye!

Over the years Project TALENT has managed to attract a fair amount of public attention and public interest, most of it favorable, and some of it from very eminent persons. For instance, the Congressional Record of January 20,

TABLE 1-1. Project TALENT original follow-up schedule

Grade When Tested in 1960	Years for Follow-up Studies			
	1-Year Follow-up	5-Year Follow-up	10-Year Follow-up	20-Year Follow-up
12	1961	1965	1970	1980
11	1962	1966	1971	1981
10	1963	1967	1972	1982
9	1964	1968	1973	1983

TABLE 1-2. Follow-ups completed thus far

Grade When Tested in 1960	Years for Follow-up Studies		
	1-year Follow-up	5-Year Follow-up	11-year Follow-up
12	1961	1965	1971
11	1962	1966	1972
10	1963	1967	1973
9	1964	1968	1974

1960, carries some very kind remarks by Senator Hubert H. Humphrey of Minnesota (later to become Vice-President Humphrey). The Congressional Record pages carrying Senator Humphrey's remarks are reproduced as Figure 1-1. Two years later, in March 1962, another Vice-President, this time the future President, Lyndon Johnson, wrote a very complimentary letter to the Project Director, John Dailey. Johnson's letter is reproduced as Figure 1-2. It may be of some interest that Lyndon Johnson and John Dailey were "noncontemporaneous schoolmates" in Southwest Texas State Teachers' College, the college in which they both earned their bachelor's degrees.

Project TALENT has also attracted its share--possibly more than its share--of unfavorable attention. In the early stages of the project it was under attack by various extremist groups, mostly at the far right edge of the political spectrum. It also attracted its share of interest in Congress. After the test battery was developed there were requests by various senators and congressmen for an opportunity to examine it. These requests, which presumably arose as a result of complaints by one or two of their constituents, were of course granted. After they examined the materials they were in every instance satisfied that whatever dire complaint had been made was without foundation, and we heard no more about it. The furthest anything like this ever got was a bill introduced in Congress early in 1959 to compel the Office of Education to cancel the contract for Project TALENT. The bill was introduced by a Representative from Michigan (who had not been among the small number of members of Congress who actually contacted us and examined the materials). Obviously the bill got nowhere. If it had been passed there would have been no occasion to write this book, seventeen years later! As Project TALENT has become better known over the years, hostility seems to have died down and support to have grown.

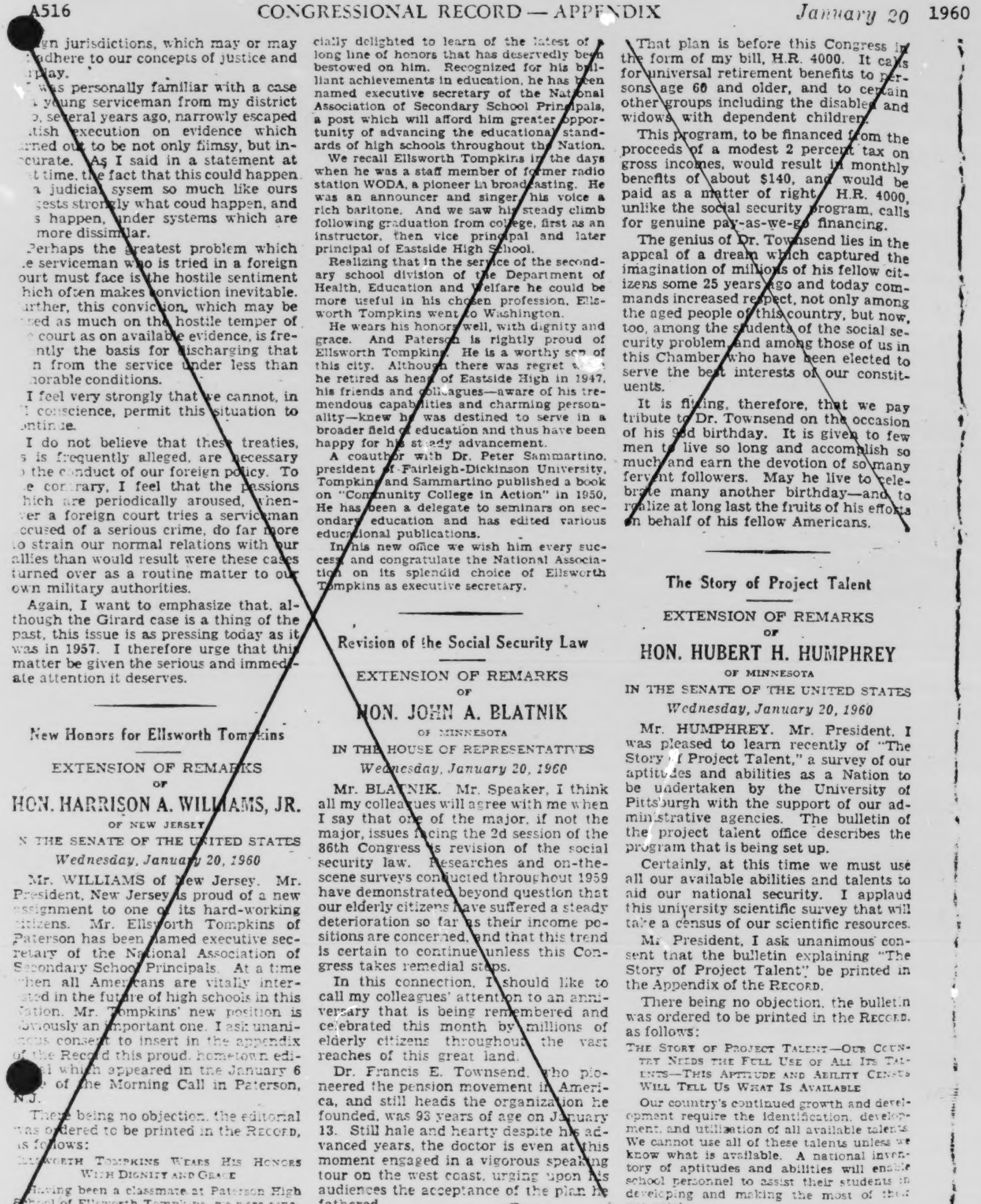


FIGURE 1-1

gn jurisdictions, which may or may adhere to our concepts of justice and play.

was personally familiar with a case young serviceman from my district o, several years ago, narrowly escaped tish execution on evidence which rned out to be not only flimsy, but in- curate. As I said in a statement at t time, the fact that this could happen a judicial sysem so much like ours gests strongly what coud happen, and s happen, under systems which are more dissimilar.

Perhaps the greatest problem which e serviceman who is tried in a foreign ourt must face is the hostile sentiment hich often makes conviction inevitable. rther, this conviction, which may be sed as much on the hostile temper of court as on available evidence, is fre- ntly the basis for discharging that n from the service under less than norable conditions.

I feel very strongly that we cannot, in conscience, permit this situation to ntinue.

I do not believe that these treaties, s is frequently alleged, are necessary the conduct of our foreign policy. To e contrary, I feel that the passions hich are periodically aroused, when- er a foreign court tries a serviceman ccused of a serious crime, do far more o strain our normal relations with our allies than would result were these cases turned over as a routine matter to our own military authorities.

Again, I want to emphasize that, although the Girard case is a thing of the past, this issue is as pressing today as it was in 1957. I therefore urge that this matter be given the serious and immediate attention it deserves.

New Honors for Ellsworth Tompkins

EXTENSION OF REMARKS

OF

HON. HARRISON A. WILLIAMS, JR.

OF NEW JERSEY

N THE SENATE OF THE UNITED STATES

Wednesday, January 20, 1960

Mr. WILLIAMS of New Jersey. Mr. President, New Jersey is proud of a new assignment to one of its hard-working citizens. Mr. Ellsworth Tompkins of Paterson has been named executive secretary of the National Association of Secondary School Principals. At a time when all Americans are vitally interested in the future of high schools in this Nation. Mr. Tompkins' new position is obviously an important one. I ask unanimous consent to insert in the appendix of the Record this proud. hometown edi- [illegible]al which appeared in the January 6 [illegible] of the Morning Call in Paterson, N.J.

There being no objection. the editorial was ordered to be printed in the Record, as follows:

Ellsworth Tompkins Wears His Honors With Dignity and Grace

Having been a classmate at Paterson High School of Ellsworth Tompkins, we were especially delighted to learn of the latest of a long line of honors that has deservedly been bestowed on him. Recognized for his brilliant achievements in education, he has been named executive secretary of the National Association of Secondary School Principals, a post which will afford him greater opportunity of advancing the educational standards of high schools throughout the Nation.

We recall Ellsworth Tompkins in the days when he was a staff member of former radio station WODA, a pioneer in broadcasting. He was an announcer and singer, his voice a rich baritone. And we saw his steady climb following graduation from college, first as an instructor, then vice principal and later principal of Eastside High School.

Realizing that in the service of the secondary school division of the Department of Health, Education and Welfare he could be more useful in his chosen profession, Ellsworth Tompkins went to Washington.

He wears his honors well, with dignity and grace. And Paterson is rightly proud of Ellsworth Tompkins. He is a worthy son of this city. Although there was regret when he retired as head of Eastside High in 1947, his friends and colleagues—aware of his tremendous capabilities and charming personality—knew he was destined to serve in a broader field of education and thus have been happy for his steady advancement.

A coauthor with Dr. Peter Sammartino, president of Fairleigh-Dickinson University, Tompkins and Sammartino published a book on "Community College in Action" in 1950. He has been a delegate to seminars on secondary education and has edited various educational publications.

In his new office we wish him every success and congratulate the National Association on its splendid choice of Ellsworth Tompkins as executive secretary.

Revision of the Social Security Law

EXTENSION OF REMARKS

OF

HON. JOHN A. BLATNIK

OF MINNESOTA

IN THE HOUSE OF REPRESENTATIVES

Wednesday, January 20, 1960

Mr. BLATNIK. Mr. Speaker, I think all my colleagues will agree with me when I say that one of the major, if not the major, issues facing the 2d session of the 86th Congress is revision of the social security law. Researches and on-the-scene surveys conducted throughout 1959 have demonstrated beyond question that our elderly citizens have suffered a steady deterioration so far as their income positions are concerned, and that this trend is certain to continue unless this Congress takes remedial steps.

In this connection, I should like to call my colleagues' attention to an anniversary that is being remembered and celebrated this month by millions of elderly citizens throughout the vast reaches of this great land.

Dr. Francis E. Townsend, who pioneered the pension movement in America, and still heads the organization he founded, was 93 years of age on January 13. Still hale and hearty despite his advanced years, the doctor is even at this moment engaged in a vigorous speaking tour on the west coast, urging upon his audiences the acceptance of the plan he fathered.

That plan is before this Congress in the form of my bill, H.R. 4000. It calls for universal retirement benefits to persons age 60 and older, and to certain other groups including the disabled and widows with dependent children.

This program, to be financed from the proceeds of a modest 2 percent tax on gross incomes, would result in monthly benefits of about $140, and would be paid as a matter of right. H.R. 4000, unlike the social security program, calls for genuine pay-as-we-go financing.

The genius of Dr. Townsend lies in the appeal of a dream which captured the imagination of millions of his fellow citizens some 25 years ago and today commands increased respect, not only among the aged people of this country, but now, too, among the students of the social security problem, and among those of us in this Chamber who have been elected to serve the best interests of our constituents.

It is fitting, therefore, that we pay tribute to Dr. Townsend on the occasion of his 93d birthday. It is given to few men to live so long and accomplish so much and earn the devotion of so many fervent followers. May he live to celebrate many another birthday—and to realize at long last the fruits of his efforts on behalf of his fellow Americans.

The Story of Project Talent

EXTENSION OF REMARKS

OF

HON. HUBERT H. HUMPHREY

OF MINNESOTA

IN THE SENATE OF THE UNITED STATES

Wednesday, January 20, 1960

Mr. HUMPHREY. Mr. President, I was pleased to learn recently of "The Story of Project Talent," a survey of our aptitudes and abilities as a Nation to be undertaken by the University of Pittsburgh with the support of our administrative agencies. The bulletin of the project talent office describes the program that is being set up.

Certainly, at this time we must use all our available abilities and talents to aid our national security. I applaud this university scientific survey that will take a census of our scientific resources.

Mr. President, I ask unanimous consent that the bulletin explaining "The Story of Project Talent" be printed in the Appendix of the Record.

There being no objection, the bulletin was ordered to be printed in the Record, as follows:

The Story of Project Talent—Our Country Needs the Full Use of All Its Talents—This Aptitude and Ability Census Will Tell Us What Is Available

Our country's continued growth and development require the identification, development, and utilization of all available talents. We cannot use all of these talents unless we know what is available. A national inventory of aptitudes and abilities will enable school personnel to assist their students in developing and making the most of their potential.

FIGURE 1-1 (continued)

If we are to plan properly for meeting the increased needs for advanced education—if we are to know, for example, how many educators or physicians we may expect to educate and train in the years ahead—we need a better idea of how many of our young people have the aptitudes, the inclination, and the education for entrance into various careers. Everyone has his own special talents. But how many persons fail to receive proper encouragement and education?

There are many tests in standard subjects, and for some years it has been possible to test one person or a thousand persons for general intelligence, musical ability, or capability in languages. Project Talent may be an historic enterprise, for it is the first scientifically planned national inventory of human talents: the aptitudes and abilities of a people.

NINETEEN SIXTY IS A CENSUS YEAR—PROJECT TALENT IS A SCIENTIFIC CENSUS

This inventory will be conducted in March of 1960, in order to coincide with the 1960 population census. When the census taker comes to the door of a home, the householder does not feel the census is an examination which is to be passed or failed. The same is true of Project Talent. No special preparation is necessary for this examination. Other sets of tests search for certain abilities or superiority in selected subjects. Project Talent will report fully the unique potential and all the talents of all students tested.

A scientifically selected sample of half a million high school students will be given a 2-day series of examinations. These tests will measure many different kinds of things. Project Talent will be carried out through local school systems all over the Nation—in the country and the city, and in public, private, and parochial schools.

The scores from the tests will be made available to the individual school, to be used as the school wishes, but Project Talent will not evaluate specific schools, nor compare on student with another. The tests will be available only to those schools selected as a part of this scientific study. The tests are not intended as substitutes for present testing programs; they are special instruments scientifically designed for this particular purpose.

This project covers more than school activities; followup studies are planned. The project will seek to determine what these people are doing in their jobs or in higher education 1 year after graduation from high school. The first survey and later questionnaires also will ask about their interests or hobbies. A young person may have a hobby which involves a special talent in photography or music, for example, but that interest may not be revealed in ordinary school work. These examinations will be designed to measure potential aptitudes and abilities of a more general sort in addition to measuring what the student knows.

Most teachers are interested not merely in school grades, but in the student's potential, and what he eventually does. When Project Talent is completed schoolteachers will have much more definite information about the kinds of talents and potential they may reasonably expect to find in their classrooms. Educators will have realistic answers to some of their questions about how to guide and develop these talents.

PROGRESS IN EDUCATIONAL RESEARCH, TEST DEVELOPMENT, AND NEW COMPUTERS MAKES POSSIBLE THIS NATIONAL INVENTORY

This national inventory of human resources has long been needed for planning in education. The scientifically selected cross section of American youth to be tested is large enough that the findings will give educators important facts about the total student population. The goals outlined below indicate some of the ways this basic information may be used.

Educators have long known that a collection of smaller studies cannot be as definitive as a national study. There are several new factors which now make such a survey possible:

Fundamental tests of aptitudes and abilities have been developed.

New electronic scoring machines, coupled with high-speed electronic computers, make it possible to process thousands of test items in a small fraction of the time required in the past. Each of the 500,000 sets of student's answer sheets will contain about 2,000 items. These 1 billion bits of information will require 1,250 machine-hours for initial scoring, reporting to schools, and recording on magnetic tape for analysis and future followup studies. The analyses of these data on high-speed computers are expected to take less than 100 hours. This data processing would have required millions of man-hours of work as recently as 5 years ago.

The principal factor in the initiation of this project, however, is that the Nation has begun to appreciate the importance of assuring an educational system designed to offer proper training and opportunities to students having diverse aptitudes and developmental needs.

Thus, leading educational research institutions, agencies of the Federal Government, and representatives of professional, educational, and scientific associations have joined together to assist in guiding the study and interpreting the results from Project Talent.

FINANCIAL SUPPORT IS BEING PROVIDED BY THE OFFICE OF EDUCATION WITH ASSISTANCE FROM OTHER FEDERAL AGENCIES

Project Talent is financed for the most part by funds from the U.S. Office of Education through Commissioner of Education Lawrence G. Derthick. These funds came from the cooperative research program, directed by Assistant Commissioner of Education Roy M. Hall. The plan for the project, as developed by the University of Pittsburgh and the American Institute for Research was reviewed and recommended for approval by the advisory committee of the cooperative research program.

The chairman of this committee was Dr. Ralph W. Tyler, Center for Advanced Study in the Behavioral Sciences, Palo Alto, Calif. Other members of the committee were Dr. Finis E. Engleman, executive secretary, American Association of School Administrators; Dr. Chester W. Harris, professor of education, University of Wisconsin; Dr. Erick L. Lindman, professor of school administration, George Peabody College for Teachers; Dr. Henry J. Otto, professor of education, University of Texas; Dr. J. Cayce Morrison, assistant commissioner for research, New York State Education Department (retired); Dr. H. H. Remmers, director, division of educational reference, Purdue University; Dr. Dewey B. Stuit, dean, College of Liberal Arts, State University of Iowa; and Dr. Dorothy Adkins Wood, chairman, department of psychology, University of North Carolina.

Support for the project is provided also by the National Institute of Mental Health and the Office of Naval Research, with aid in the planning stages from the National Science Foundation.

Dr John C. Flanagan, professor at the University of Pittsburgh and director of the American Institute for Research, prepared the initial proposals for the study and is the responsible investigator.

THE PLANNING AND ADMINISTRATION OF PROJECT TALENT IS UNDER THE GUIDANCE OF NATIONAL COMMITTEES AND COORDINATORS

Four panels of the Nation's leading specialists worked with the staff for more than a year planning the study. This staff included six full-time research workers headed by Dr. John T. Dailey. The overall chairman of the advisory panels during the planning phases was Dr. John H. Fischer, then superintendent of the Baltimore Public Schools. The present chairman is Dr. Kenneth E. Oberholtzer, superintendent of the Denver public schools. The chairman of the guidance and counseling panel is Dr. Edward Landy, director, division of counseling services, Newton (Mass.) public schools. The chairman of the educational research panel is Dr. Robert J. Keller, director, University High School, University of Minnesota. The chairman of the testing problems panel is Dr. Robert L. Thorndike, head, department of psychological foundations and services, Teachers College, Columbia University. The chairman of the manpower and sociology panel is Dr. Samuel A. Stouffer, director, laboratory of social relations, Harvard University.

The plans developed through these groups were reviewed by an advisory council composed principally of presidents and executive secretaries of the professional associations most interested in the findings from this study. These associations included the American Association of School Administrators, National Association of Secondary-School Principals, Council of Chief of School Officers, National Catholic Education Association, National School Boards Association, American Association for the Advancement of Science, American Personnel and Guidance Association, and American Psychological Association.

Regional coordinators will administer the program locally. They will distribute tests, arrange for cooperation of schools, and serve as advisers on procedures. These local officials, familiar with the educational needs and resources of the local communities, will be the key representatives of one of the largest educational studies ever undertaken.

All results from the survey and the studies which follow will be published and will be freely available to educators and research specialists. Prior to publication, these studies will be reviewed by the panel of experts and the advisory council. Many educational and research groups have expressed keen interest in these studies and will be kept fully informed. A series of reports will be made to educational agencies throughout the country. These will go to school systems, to schools of education, to State departments of education, and to private educational associations as well as to the supporting agencies. Results will be of interest also to parents, to industry, to the military services, and to anyone concerned with planning for proper future use of manpower.

The responsibility for administering the tests and collecting the data about the students and their past experiences is placed directly in the hands of the selected schools. The scientific analysis of the original data and the collection of the follow-up data will be done under the direction of the program staff of the University of Pittsburgh with special assistance from the American Institute for Research.

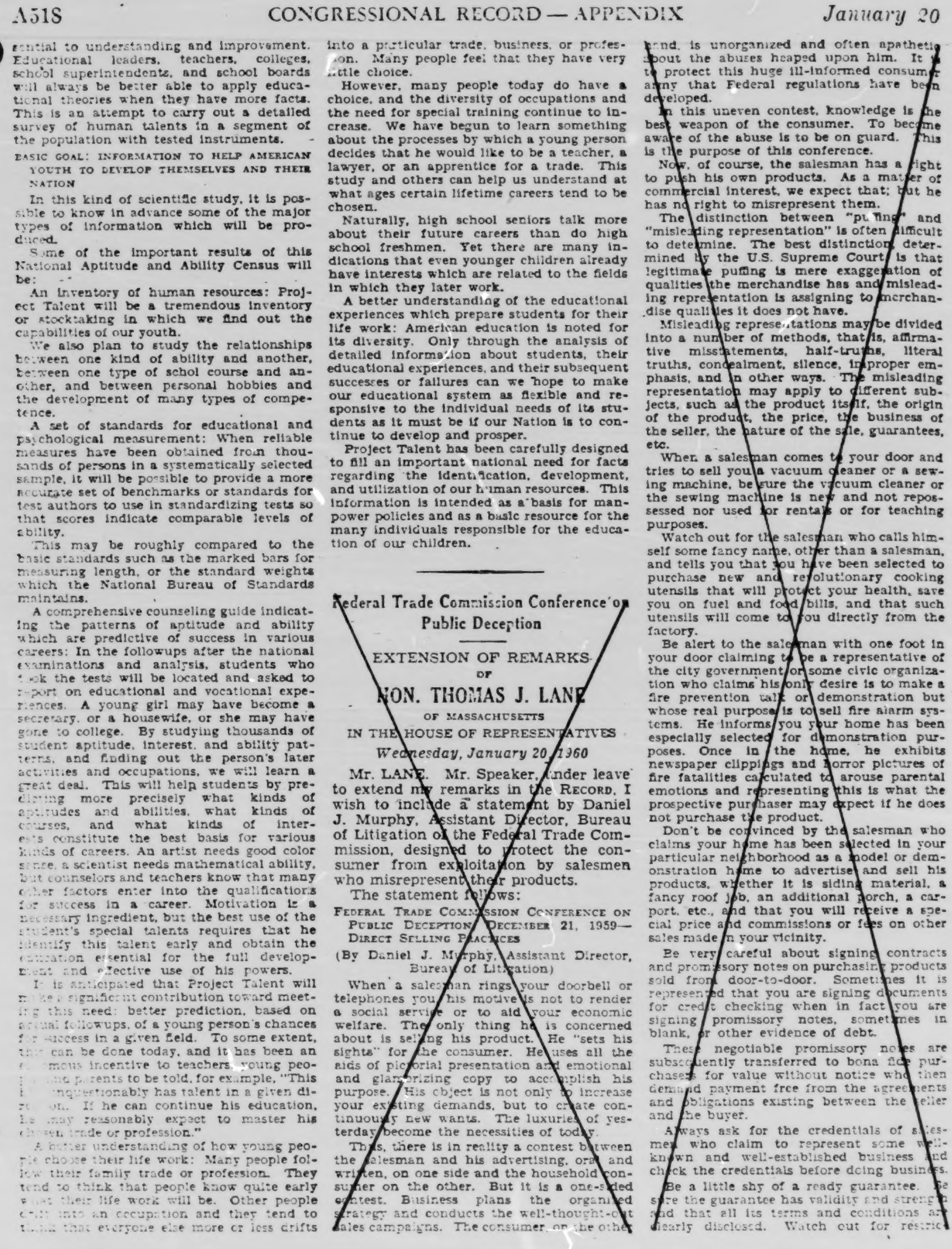
The success of the survey phase depends upon the capability and effective assistance of local coordinators, principals, teachers, and others.

PROJECT TALENT IS A SCIENTIFIC STUDY USING ONLY TESTED TYPES OF INSTRUMENTS

It is expected that this historic national census and follow-up will give the Nation a great many new facts. Some will be surprising. Also, it is to be expected that administration of such a large program will produce new methods and new ideas in the field of educational research itself. This should be one of several important by-products from the study.

"Project Talent" is not an experiment to try out new theories. All the major components of this project have been tested and proven in previous studies. Systematic measurement and description of students is es-

FIGURE 1-1 (continued)



sential to understanding and improvement. Educational leaders, teachers, colleges, school superintendents, and school boards will always be better able to apply educational theories when they have more facts. This is an attempt to carry out a detailed survey of human talents in a segment of the population with tested instruments.

BASIC GOAL: INFORMATION TO HELP AMERICAN YOUTH TO DEVELOP THEMSELVES AND THEIR NATION

In this kind of scientific study, it is possible to know in advance some of the major types of information which will be produced.

Some of the important results of this National Aptitude and Ability Census will be:

An inventory of human resources: Project Talent will be a tremendous inventory or stocktaking in which we find out the capabilities of our youth.

We also plan to study the relationships between one kind of ability and another, between one type of schol course and another, and between personal hobbies and the development of many types of competence.

A set of standards for educational and psychological measurement: When reliable measures have been obtained from thousands of persons in a systematically selected sample, it will be possible to provide a more accurate set of benchmarks or standards for test authors to use in standardizing tests so that scores indicate comparable levels of ability.

This may be roughly compared to the basic standards such as the marked bars for measuring length, or the standard weights which the National Bureau of Standards maintains.

A comprehensive counseling guide indicating the patterns of aptitude and ability which are predictive of success in various careers: In the followups after the national examinations and analysis, students who took the tests will be located and asked to report on educational and vocational experiences. A young girl may have become a secretary, or a housewife, or she may have gone to college. By studying thousands of student aptitude, interest, and ability patterns, and finding out the person's later activities and occupations, we will learn a great deal. This will help students by predicting more precisely what kinds of aptitudes and abilities, what kinds of courses, and what kinds of interests constitute the best basis for various kinds of careers. An artist needs good color sense, a scientist needs mathematical ability, but counselors and teachers know that many other factors enter into the qualifications for success in a career. Motivation is a necessary ingredient, but the best use of the student's special talents requires that he identify this talent early and obtain the education essential for the full development and effective use of his powers.

It is anticipated that Project Talent will make a significant contribution toward meeting this need: better prediction, based on actual followups, of a young person's chances for success in a given field. To some extent, this can be done today, and it has been an enormous incentive to teachers, young people, and parents to be told, for example, "This boy unquestionably has talent in a given direction. If he can continue his education, he may reasonably expect to master his chosen trade or profession."

A better understanding of how young people choose their life work: Many people follow their family trade or profession. They tend to think that people know quite early what their life work will be. Other people drift into an occupation and they tend to think that everyone else more or less drifts into a particular trade, business, or profession. Many people feel that they have very little choice.

However, many people today do have a choice, and the diversity of occupations and the need for special training continue to increase. We have begun to learn something about the processes by which a young person decides that he would like to be a teacher, a lawyer, or an apprentice for a trade. This study and others can help us understand at what ages certain lifetime careers tend to be chosen.

Naturally, high school seniors talk more about their future careers than do high school freshmen. Yet there are many indications that even younger children already have interests which are related to the fields in which they later work.

A better understanding of the educational experiences which prepare students for their life work: American education is noted for its diversity. Only through the analysis of detailed information about students, their educational experiences, and their subsequent successes or failures can we hope to make our educational system as flexible and responsive to the individual needs of its students as it must be if our Nation is to continue to develop and prosper.

Project Talent has been carefully designed to fill an important national need for facts regarding the identification, development, and utilization of our human resources. This information is intended as a basis for manpower policies and as a basic resource for the many individuals responsible for the education of our children.

~~Federal Trade Commission Conference on Public Deception~~

~~EXTENSION OF REMARKS~~

~~OF~~

~~HON. THOMAS J. LANE~~

~~OF MASSACHUSETTS~~

~~IN THE HOUSE OF REPRESENTATIVES~~

~~*Wednesday, January 20, 1960*~~

~~Mr. LANE. Mr. Speaker, under leave to extend my remarks in the RECORD, I wish to include a statement by Daniel J. Murphy, Assistant Director, Bureau of Litigation of the Federal Trade Commission, designed to protect the consumer from exploitation by salesmen who misrepresent their products.~~

~~The statement follows:~~

~~FEDERAL TRADE COMMISSION CONFERENCE ON PUBLIC DECEPTION, DECEMBER 21, 1959—DIRECT SELLING PRACTICES~~

~~(By Daniel J. Murphy, Assistant Director, Bureau of Litigation)~~

~~When a salesman rings your doorbell or telephones you, his motive is not to render a social service or to aid your economic welfare. The only thing he is concerned about is selling his product. He "sets his sights" for the consumer. He uses all the aids of pictorial presentation and emotional and glamorizing copy to accomplish his purpose. His object is not only to increase your existing demands, but to create continuously new wants. The luxuries of yesterday become the necessities of today.~~

~~Thus, there is in reality a contest between the salesman and his advertising, oral and written, on one side and the household consumer on the other. But it is a one-sided contest. Business plans the organized strategy and conducts the well-thought-out sales campaigns. The consumer, on the other hand, is unorganized and often apathetic about the abuses heaped upon him. It is to protect this huge ill-informed consumer army that Federal regulations have been developed.~~

~~In this uneven contest, knowledge is the best weapon of the consumer. To become aware of the abuse is to be on guard. This is the purpose of this conference.~~

~~Now, of course, the salesman has a right to push his own products. As a matter of commercial interest, we expect that; but he has no right to misrepresent them.~~

~~The distinction between "puffing" and "misleading representation" is often difficult to determine. The best distinction determined by the U.S. Supreme Court is that legitimate puffing is mere exaggeration of qualities the merchandise has and misleading representation is assigning to merchandise qualities it does not have.~~

~~Misleading representations may be divided into a number of methods, that is, affirmative misstatements, half-truths, literal truths, concealment, silence, improper emphasis, and in other ways. The misleading representation may apply to different subjects, such as the product itself, the origin of the product, the price, the business of the seller, the nature of the sale, guarantees, etc.~~

~~When a salesman comes to your door and tries to sell you a vacuum cleaner or a sewing machine, be sure the vacuum cleaner or the sewing machine is new and not repossessed nor used for rentals or for teaching purposes.~~

~~Watch out for the salesman who calls himself some fancy name, other than a salesman, and tells you that you have been selected to purchase new and revolutionary cooking utensils that will protect your health, save you on fuel and food bills, and that such utensils will come to you directly from the factory.~~

~~Be alert to the salesman with one foot in your door claiming to be a representative of the city government or some civic organization who claims his only desire is to make a fire prevention talk or demonstration but whose real purpose is to sell fire alarm systems. He informs you your home has been especially selected for demonstration purposes. Once in the home, he exhibits newspaper clippings and horror pictures of fire fatalities calculated to arouse parental emotions and representing this is what the prospective purchaser may expect if he does not purchase the product.~~

~~Don't be convinced by the salesman who claims your home has been selected in your particular neighborhood as a model or demonstration home to advertise and sell his products, whether it is siding material, a fancy roof job, an additional porch, a carport, etc., and that you will receive a special price and commissions or fees on other sales made in your vicinity.~~

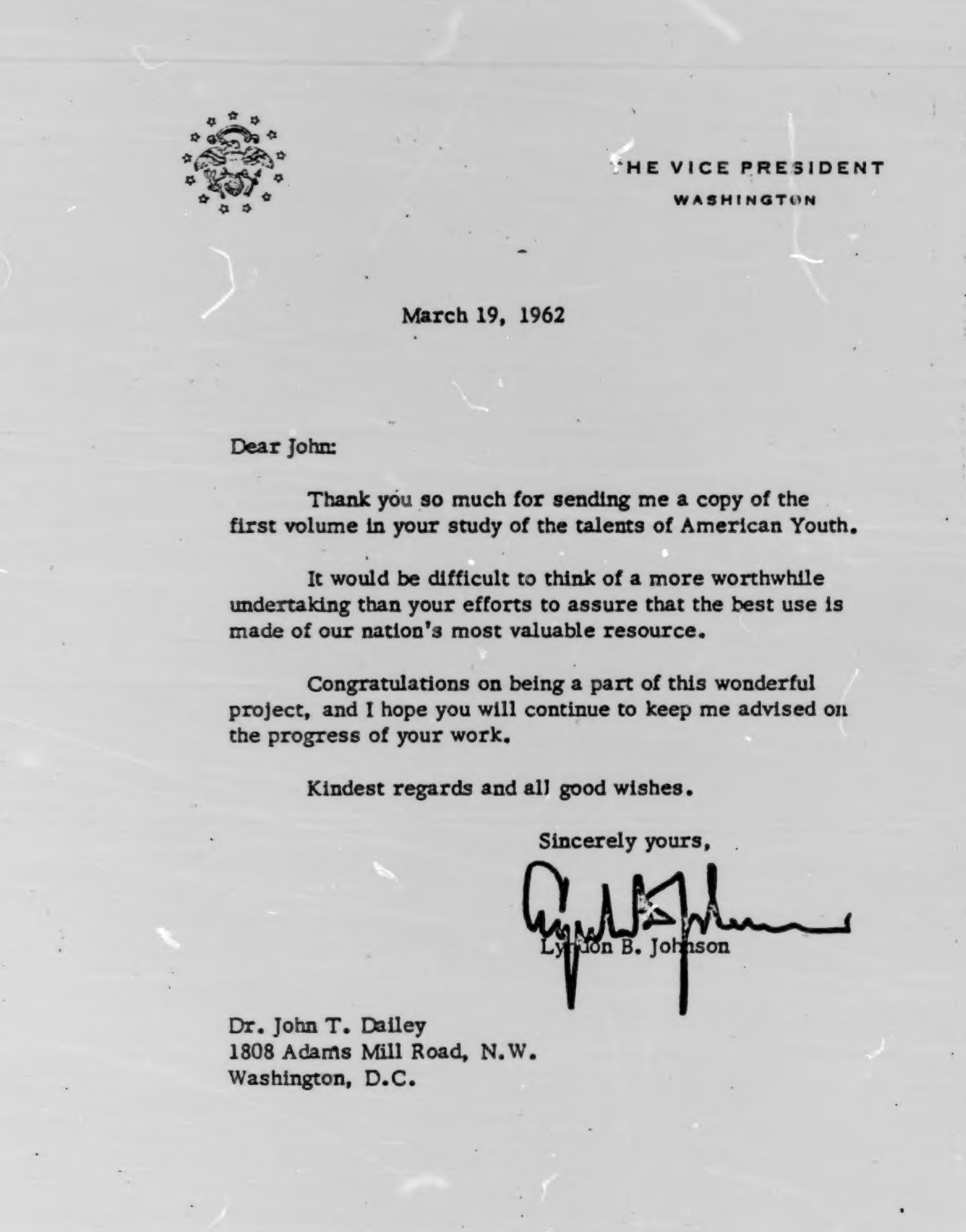
~~Be very careful about signing contracts and promissory notes on purchasing products sold from door-to-door. Sometimes it is represented that you are signing documents for credit checking when in fact you are signing promissory notes, sometimes in blank, or other evidence of debt.~~

~~These negotiable promissory notes are subsequently transferred to bona fide purchasers for value without notice who then demand payment free from the agreements and obligations existing between the seller and the buyer.~~

~~Always ask for the credentials of salesmen who claim to represent some well-known and well-established business and check the credentials before doing business.~~

~~Be a little shy of a ready guarantee. Be sure the guarantee has validity and strength and that all its terms and conditions are clearly disclosed. Watch out for restric~~

FIGURE 1-2



THE VICE PRESIDENT
WASHINGTON

March 19, 1962

Dear John:

Thank you so much for sending me a copy of the first volume in your study of the talents of American Youth.

It would be difficult to think of a more worthwhile undertaking than your efforts to assure that the best use is made of our nation's most valuable resource.

Congratulations on being a part of this wonderful project, and I hope you will continue to keep me advised on the progress of your work.

Kindest regards and all good wishes.

Sincerely yours,

Lyndon B. Johnson

Dr. John T. Dailey
1808 Adams Mill Road, N.W.
Washington, D.C.

Chapter 2: THE PRENATAL STAGE:
(Project TALENT Before It Was Project TALENT)

A. Roots of the idea

The initial idea for Project TALENT goes back as far as December 1956. At that time John Flanagan was working on the technical manual for the Flanagan Aptitude Classification Tests (FACT Battery) and was concerned about the problem of norms. Data were available on about 3,000 or 4,000 cases but he didn't regard this as a really adequate sample for the norms. Pondering the thought that it was very difficult (perhaps impossible) for the individual test-developer to arrange for a really satisfactory norm group, he decided that the only feasible solution would be for the Federal government to do it. What would be needed would be a very large and truly representative sample of the population to which the norms were to apply. If a fairly comprehensive "anchor battery" of tests were administered to such a sample it would be to the advantage of all test authors and all test publishers, since through the medium of equipercentile equating it would then be possible to develop excellent norms for any test that was comparable (in level and type of content) to a test in the anchor battery. Checking back in the literature to see whether anything like this idea had been suggested he found that over a quarter of a century earlier Herbert Toops had suggested something similar. Toops had proposed that a million-member sample be selected, representative of the entire American population; he dubbed this proposed sample the "Standard Million". This sample would be used for the standardization of all sorts of tests.

Because he was focusing primarily on the high school level John Flanagan was able to modify the Toops idea by making it a bit less general. The sample would still be representative of a huge segment of the American population, but not of the entire population. The population represented would consist of everybody in high school at a given time. Because of this restriction, the size of the sample could be scaled down, from a million to a mere (?) half-million. That meant that roughly 5 percent of all high school students would be in the sample. A considerably larger proportion of the corresponding population than Toops' "Standard Million" would be represented; that would have been only about three-quarters of a percent.

If Toops' Standard Million was one root of the idea for Project TALENT, what were the others? The answer to that question lies in large part in the history of aptitude and achievement testing and of research with the development and validation of methods for predicting success in various careers. John Flanagan had played a major role in these areas; in his capacity as Associate Director of the Cooperative Test Service (CTS) of the American Council on Education he had played a leading role in the development and scaling of achievement tests. And in his World War II role as the Director of the Army Air Force Aviation Psychology Program he blazed some trails in the development of aptitude tests and in differential prediction (the basic problem in personnel classification).

In developing plans for the project, John was of course fully cognizant at all stages not only of the tremendous possibilities the project would offer but also of the difficult problems and potential pitfalls that would be encountered. For instance he knew that it would be necessary to cope with the criterion problem, a stumbling block which had been the downfall of so many previous research projects. In an address given at Lake Washington High School for the Counseling and Guidance Training Institute, University of Washington, on April 21, 1960, he said:

> "...How is it planned to determine 'the specific patterns of aptitudes and abilities and interests which provide the best basis for college courses and careers'? We plan to do this by comparing the thousand items of information about these people with their subsequent success in school courses, college courses and occupational experiences. We

> will do this by following these students up 1 year, 5 years, 10 years, and 20 years after they graduate from high school. We do not plan to retest them all at this time but we will use follow-up questionnaires to get information as to what courses they're taking, what grades they are getting in their college courses, what employment they have been in, what salaries, how satisfied, how successful they are, and similar types of information. We hope that before 20 years have gone by we will be able to make some improvements on methods for assessing the success of people in various occupations such as teaching so that we will be able to do some comparison within groups."

In that last sentence Flanagan was referring to the by then well-known fact that it is much easier to validate predictive tests in terms of how well they differentiate among groups—in other words validating against the group-membership criterion—than in terms of how well they correlate with levels of success within group. The last sentence in the quotation above mentions a 20-year goal for solving this difficult problem. If that goal is to be met there are only about 4 years left. In the 16 years that have elapsed since then some thought has been given to the problem and perhaps some progress has been made, but the problem is by no means solved.

Flanagan's speech continues as follows, pointing out some additional areas where problems could be anticipated, and indicating how Project TALENT might bypass them:

> "Studies such as 'Ten Thousand Careers'* and similar follow-up studies have not been very successful in differentiating the people who are most successful in occupations from the people who are less successful. But many of these studies have had special problems. For example, in Thorndike's 'Ten Thousand Careers', the group was quite heterogeneous in the amounts of education and age. This group of people came into

---

* At about the time that the plans for what was to become Project TALENT were first being developed (1957) Robert L. Thorndike and Elizabeth Hagen were engaged in a monumental follow-up study of about 10,000 men who had been in the Air Force during World War II. The results are reported in a most enlightening book, entitled "Ten Thousand Careers" (1959).

> the Air Force, some after completing college and some right out of high school. When Thorndike followed them up 12 years later, some of them had just recently finished their schooling while others had ten or more years of experience in a field after college. We hope that members of our group are sufficiently similar with respect to age and to educational experience that we will be able to make comparisons without some of the disturbing factors which made it difficult to establish relations in the Thorndike data. We still recognize, however, that the factors that cause one lawyer to earn $100,000 and another lawyer to earn only $5,000* a year may not be predictable from any of the present kinds of tests."

In this section we have discussed, on a somewhat limited basis, the roots of the idea for Project TALENT. From those roots the full-scale idea began to burgeon, including the principal characteristics that would define its scope: large-scale testing, on an enormous sample, with a very comprehensive battery of objective aptitude and achievement tests and questionnaires, long-term follow-up, and predictive validation.

B. The start of the search for funding

Clearly the idea that was eventually to develop into "Project TALENT" was a very ambitious one--indeed, an audacious one--and it would take a very large amount of money, over a long period of time, to put it into effect.

How does one go about getting funding for an educational research project that one foresees is likely to cost seven or eight million dollars before it is through? (Actually the project total cost is now substantially higher than that, because of the inflation that has occurred since 1957.) And in particular, how can one hope to accomplish this in a milieu not geared to large-scale costly educational research projects? Remember, back in the late 1950's the National Institute of Education didn't exist, and the Office of Education was a much smaller-scale operation than it is now. Educational research projects much more limited in scope but costing nearly as much (or even, in some cases, more) are now almost commonplace, but in 1957 the idea was daring in its magnitude. To repeat the question posed at the beginning of this paragraph: How, in 1957, would one go about getting funding for such a project? The answer, if the person seeking the funding happens to be John Clemans Flanagan, can probably be summed up in two words: indefatigability and thoroughness.

---

* In 1960, levels of earnings and cost of living were far lower than now!

He immediately wrote a prospectus for the project and discussed it informally with the heads of many Federal government agencies--including, among others, the National Science Foundation (NSF), National Institutes of Health (NIH), Office of Education (OE), National Institute of Mental Health (NIMH), Office of Naval Research (ONR), Research Division of the Bureau of Naval Personnel (BuPers), Personnel Research Branch (PRB) of the Adjutant General's Office of the Department of the Army, and the Personnel Laboratory of the Air Force Personnel and Training Research Center (AFPTRC). The response was generally favorable in the sense that the government personnel with whom the idea was discussed thought it was a great idea and that someone ought to fund it. None of them, however, perceived his own agency as being the appropriate one to finance the project.

At this early stage the plan for the project was almost identical with the plan that ultimately developed. Most of the main features were there: the large national sample consisting of all students in grades 9-12 in a random sample of all high schools, the aptitude testing with a comprehensive battery, the comprehensive biographical inventory blank, and the follow-ups by questionnaire 1, 5, 10, and 20 years after high school graduation. There was one major difference, however. Since it was the problem of developing norms for the FACT Battery that had been the source of the idea for the project in the first place, and since the FACT battery comprised a very comprehensive set of aptitude tests which would be eminently suitable for the purposes of the project, naturally the use of that battery was an inherent part of the original plan. (As Chapter 3 indicates, however, it turned out that there were excellent reasons for changing that part of the plan in favor of developing a new battery of tests designed especially for the project; eventually, therefore the part of the plan that involved use of the FACT Battery was completely changed.)

After the groundwork had been laid by innumerable informal conferences at which the preliminary prospectus was discussed, the next step was to organize a planning conference, the purpose of which would be to give agencies planning to participate in the program or having a direct interest in its outcomes an opportunity to discuss the design of the study and contribute their ideas prior to the preparation of a formal proposal. The organizations invited to send representatives to the planning conference, which was held late in February 1957, included government agencies, foundations, professional organizations such as the American Psychological Association (APA), the American Association for the Advancement of Science (AAAS), and the National Education Association (NEA), and test publishers. It was hoped that if project plans could be developed in such a way that the needs of all these various organizations would be served, there might result a groundswell of support for the project.

Among those who attended the meeting was Alice Scates of the Office of Education. She was later to become Project TALENT's first monitor. The meeting was generally successful, in that a favorable attitude toward the prospective project was generated--but there was still not one cent of funding in hand.

To permit plans to be developed, the Cooperative Research Program of the Office of Education agreed to provide a five-year grant (July 1957 - June 1962) of about $114,000 to cover project planning, "get ready costs," and some other miscellaneous expenses. Because the Cooperative Research Program was by law not permitted to make grants to any kind of organization except state departments of education and institutions of higher learning, the grant could not go directly to AIR. Instead, the proposal written by John Flanagan was submitted by the University of Pittsburgh, where he was Professor of Psychology; and it was to the University of Pittsburgh that the Cooperative Research Program awarded the grant.

The Principal Investigator, who was of course John Flanagan, appointed as the incipient project's first Project Director Robert C. Craig, who was at that time a member of AIR's research staff. (A little over a year later, in November 1958, the Cooperative Research Program awarded another grant to Pitt. This one was for $150,000 and was to run for 2 years; it was for a supplementary sample of 16-year-olds not in grade 9, 10, 11, or 12.) Meanwhile early in 1958 NIMH provided a supplementary grant ($46,000) to the University of Pittsburgh, and ONR and NSF awarded small contracts to AIR ($35,000 and $15,000 respectively) to provide supplementary support for the project.

Thus at this point a total of $360,000 altogether had been obtained (mostly from the Office of Education, but partly from three other agencies). This $360,000 welcome though it was, was certainly not enough to assure the future of the project. The efforts to get funding for going beyond the planning stage, by actually carrying the project out, could not and did not abate. Prospectuses and proposals continued to be written, discussed, revised, submitted.

One very important step that took place early in 1958 was the appointment of four Advisory Panels. They were in the following areas and had the following chairmen:

	Chairman
1. Guidance and Counseling	Edward Landy, Director, Division of Counseling Services, Newton (Massachusetts) Public Schools
2. Educational Research	Robert J. Keller, Director, University High School, University of Minnesota
3. Testing Problems	Robert L. Thorndike, Head, Department of Psychological Foundations and Services, Teachers College, Columbia University
4. Manpower and Sociology	Samuel A. Stouffer, Director, Laboratory of Social Relations, Harvard University

Membership of the Panels is shown in Figure 2-1.

It will be noted that the first overall chairman of the Panels was John Fischer, at that time Baltimore's School Superintendent. Originally Ralph Tyler, the Director of the Center for Advance Study in the Behavioral Sciences in Palo Alto, had been slated to be Chairman, but he had resigned from this position before any meetings were held, because he perceived that in view of his appointment as Chairman of the OE Cooperative Research Program's Research Advisory Committee for 1958-59, there might be a conflict of interest. The Research Advisory Committee is shown in Figure 2-2. When Dr. Tyler therefore resigned as Chairman of the Advisory Panels, Dr. Fischer agreed to become his replacement. In the spring of 1959, Dr. Fischer, too, resigned, because he was leaving his position as Baltimore Superintendent of Schools to become the Dean of Teachers College, Columbia University. (He is now the President of that institution and has been since 1962.) Dr. Kenneth Oberholtzer, Denver's School Superintendent, replaced Dr. Fischer as Chairman.

FIGURE 2-1. Advisory Panels

Chairman (seriatim):

John H. Fischer, Superintendent, Baltimore Public Schools
Kenneth Oberholtzer, Superintendent of the Denver Public Schools

GUIDANCE AND COUNSELING PANEL

Chairman, Edward Landy, Director, Division of Counseling Services, Newton (Mass.) Public Schools

Ralph F. Berdie, Director, Student Counseling Bureau, University of Minnesota

Bruce E. Shear, Chief, Bureau of Guidance, New York State Education Department

John M. Stalnaker, President, National Merit Scholarship Corporation

David V. Tiedeman, Associate Professor of Education, Harvard University

Arthur E. Traxler, Executive Director, Educational Records Bureau

Leona E. Tyler, Professor of Psychology, University of Oregon

EDUCATIONAL RESEARCH PANEL

Chairman, Robert J. Keller, Director, University High School, University of Minnesota

Reverend O'Neil C. D'Amour, Assistant to the Executive Secretary, National Catholic Education Association

Paul L. Banfield, Headmaster, Landon School (Maryland)

Warren G. Findley, Assistant Superindendent for Pupil Personnel Services, Atlanta Public Schools

Earl J. McGrath, Professor of Education, Teachers College, Columbia University

Paul T. Rankin, Assistant Superintendent of Schools, Detroit Public Schools

James W. Reynolds, Professor of Education, University of Texas

J. Wayne Wrightstone, Director, Bureau of Educational Research, New York City Board of Education

TESTING PROBLEMS PANEL

Chairman, Robert L. Thorndike, Head, Department of Psychological Foundations and Services, Teachers College, Columbia University

Henry Chauncey, President, Educational Testing Service

Wayne H. Holtzman, Associate Director, Hogg Foundation, University of Texas

A. Paul Horst, Professor of Psychology, University of Washington

Lloyd G. Humphreys, Chairman, Department of Psychology, University of Illinois

E. Lowell Kelly, Director, Bureau of Psychological Services, University of Michigan

Joseph Zubin, Principal Research Scientist, Biometrics Research Unit, New York State Department of Mental Hygiene

MANPOWER AND SOCIOLOGY PANEL

Chairman, Samuel A. Stouffer, Director, Laboratory of Social Relations, Harvard University

E. Franklin Frazier, Head, Department of Sociology, Howard University

Seymour E. Harris, Chairman, Department of Economics, Harvard University

Donald G. Marquis, Massachusetts Institute of Technology

Irving D. Lorge, Executive Officer, Institute of Psychological Research, Teachers College, Columbia University

C. Joseph Nuesse, Dean, School of Social Science, Catholic University of America

Fred L. Strodtbeck, Associate Professor of Sociology, University of Chicago

FIGURE 2-2. Research Advisory Committee of
The Cooperative Research Program (1958-59)

U.S. Office of Education

Commissioner Lawrence G. Derthick

Assistant Commissioner Roy M. Hall

Committee Members

Chairman, Ralph W. Tyler, Director
Center for Advanced Study in the Behavioral Sciences
Stanford University

Finis E. Engleman, Executive Secretary
American Association of School Administrators

Chester W. Harris, Professor of Education
University of Wisconsin

Erick L. Lindman, Professor of School Administration
George Peabody College for Teachers

Henry J. Otto, Professor of Education
University of Texas

J. Cayce Morrison, Assistant Commissioner for Research
New York State Education Department (retired)

H. H. Remmers, Director
Division of Educational Reference, Purdue University

Dewey B. Stuit, Dean
College of Liberal Arts, State University of Iowa

Dorothy Adkins Wood, Chairman
Department of Psychology, University of North Carolina

After the four Advisory Panels had met quite a few times, an Advisory Council was set up; its principal function was to review plans developed by the Advisory Panels. The Advisory Council was composed mostly of Presidents and Executive Secretaries of the professional associations most interested in the project and its findings. Membership of the Council is shown in Figure 2-3.

One result of the existence of this Council was that in effect it provided a channel of communication between the Principal Investigator and the organizations represented by the Council members. Factions in some of these organizations had been opposed to the project because they were afraid it would result in unfair and inappropriate comparisons of different schools, or that unwarranted inferences might be made about the relative merits of the school systems in different states, merely on the basis of a comparison of the results for a handful of schools. For the most part it was possible to allay this uneasiness by making a firm commitment that not only would no data be released concerning individuals but that we would go far beyond that, in that we would not release data about any participating school, or school system, or city, or county, or state. In fact the smallest geographical unit about which our commitment _would_ permit us to release data would be _region_ of the country. Furthermore we made a commitment not to release data that would permit comparisons among the three major categories of school in terms of type of control (i.e., public, parochial, and other private schools).

C. The planning and the plans

As the basic plans took shape ever more firmly, it became apparent that the eventual design of the project would be very similar to John Flanagan's early idea, which had first popped into his head in December 1956. The following features were agreed upon early by all concerned (although not necessarily in the following words):

1. It would be a survey study, not an experimental study.

FIGURE 2-3. Advisory Council

Martin Essex, President
American Association of School Administrators

George E. Watson, President
Council of Chief State School Officers

Edgar Fuller, Executive Secretary
Council of Chief State School Officers

Dugald S. Arbuckle, President
American Personnel and Guidance Association

Arthur A. Hitchcock, Executive Secretary
American Personnel and Guidance Association

Paul E. Klopsteg, President
American Association for the Advancement of Science

Dael Wolfle, Administrative Secretary
American Association for the Advancement of Science

Donald O. Hebb, President
American Psychological Association

John G. Darley, Executive Secretary
American Psychological Association

Robert E. Willis, President
National School Boards Association

W. A. Shannon, Executive Director
National School Boards Association

Cliff Robinson, President
National Association of Secondary School Principals

The Rt. Rev. Msgr. T. Leo Keaveny, President
Secondary School Department
National Catholic Education Association

The Rt. Rev. Msgr. F. G. Hochwald, Executive Secretary
National Catholic Education Association

2. The sample would be very large, and it would be a true probability sample, selected by random means.
3. The study would be both cross-sectional and longitudinal.
4. Its longitudinal feature would be a long-term one, with several follow-ups, over as long a period as 20 years.

The panels met quite a few times. There were two two-day meetings in the spring of 1958, one the following fall, and another the following spring. There was also considerable consultation by mail. Thus the panels were very active ones. (Further details about what was decided on the basis of their recommendations are provided in the next chapter.)

D. Finding funding

Recognizing that it was in the public interest to get the project funded, John Wilson, Assistant Director of the National Science Foundation, took the lead in convening a Federal Inter-Agency Committee to consider the problem. The committee met late in 1958. The organizations represented included the Army, Navy, Air Force, National Science Foundation, Atomic Energy Commission, National Institutes of Health, and, of course, the Office of Education. John Flanagan of course attended, too. Though all the participating agencies were interested in the project and wanted to see it get under way, none had been ready to fund it. In the course of the committee deliberations, however, a consensus was reached that in view of the nature of the proposed project, by far the most appropriate agency to fund it would be the organization that was already providing the principal support for the planning phase--the Office of Education. Roy M. Hall, the Assistant Commissioner of Education, who was a participant in the Interagency Committee meeting, agreed. And since he happened to be Director of the Cooperative Research Program, he was in a position to recommend that that agency finance the project. Fortunately the Commissioner of Education, Lawrence G. Derthick, agreed with him. Thus some very important hurdles had been passed in the search for funds. The proposal was formally submitted to the Office of Education in December 1958. But despite the backing of Dr. Derthick and Dr. Hall, one major hurdle yet remained: all projects to be funded by the Cooperative Research Program had to be approved first by the Research Advisory Committee. (See Figure 2-2.) Fortunately, under the chairmanship of Ralph Tyler, who understood the importance of the project and its potential value, the Committee gave its blessing! That was in the spring of 1959. In April 1959 the Cooperative Research Program awarded a $480,000 grant to the University of Pittsburgh, and Project TALENT was at last on its way! The prenatal stage was over. It had been about the same length (28 months) from conception in December 1956 to birth in April 1959 as the gestation period of an elephant--a coincidence that probably has no significance whatever.

The University of Pittsburgh, not AIR, was of course the prime grantee for the same reason that applied in the case of the planning grants. However, it was fully understood by all concerned that the project was to be carried out as a cooperative venture of the University and AIR, with the University's contribution being mainly to provide administrative services. All AIR staff members who were to be on the Project TALENT staff received University appointments (while continuing also to be associated with AIR). These arrangements, while they caused a little awkwardness and inconvenience, nevertheless worked quite well.

E. Bonfires and brouhahas

We did not start to develop the Project TALENT Battery until about the beginning of 1959. But in the meantime we were busy with experimental tryouts of pseudo-batteries composed of parts of commercial tests; in some instances, after preliminary forms of our own instruments had been developed, we arranged to incorporate them in other researchers' batteries for a single administration. The purposes of these early tryouts varied. In some cases the aim was to work out the logistical and other details of handling the administration of a very large test battery to all students in a school. Another purpose was to get some preliminary data (intercorrelations, for instance) that might prove helpful in our test development efforts. A third purpose, at least later in the 1958-59 school year, when development of preliminary forms of most of our tests and questionnaires was well under way, was to get some empirical data on the new items--some preliminary item analysis data. In two of the communities where activities of this sort were undertaken, personality inventories were being used that had not been constructed by TALENT personnel, were not intended to be part of the TALENT Battery, and were in both instances somebody else's instruments--NOT OURS! But there was sufficient confusion that when a few outraged parents took umbrage at some of the personality inventory questions that they regarded as an offensive invasion of privacy, Project TALENT was blamed, and for a while the press was enlivened by reports of public protests against the project, defense of the project by school personnel, and liberal quoting of the items which aroused the ire of those parents and other members of the public who were most vociferous. For convenience, we may name these two episodes after the sites where they occurred--calling them respectively "the Baltimore brouhaha" and "the Texas bonfire." Some readers of this

History may remember reading in their local newspapers something about one or both of these episodes at the time they occurred. Whey they read is very likely to have been wrong--either in toto or in large part. Let us try to set the record straight.

1. The Baltimore brouhaha

Arrangements were made with a school in Reistertown, Maryland, a small community in Baltimore County, to administer a "pseudo-battery"* to the students. Among the instruments included was the Minnesota Counseling Inventory, a widely used personality questionnaire published by the Psychological Corporation. Three items in this questionnaire angered some parents and others who heard about them and unfortunately attributed them to Project TALENT. The items that led to the resulting brouhaha were:

1. Do you dream about sex?
2. I am worried about sex.
3. Have you ever stolen anything?

The Baltimore newspapers had articles on it for several days. The matter was discussed at PTA meetings and at faculty meetings in the high school in question. School authorities stalwartly defended the Minnesota Counseling Inventory and Project TALENT. (They did not distinguish between the two.) The school faculty passed the following resolution:

> "Resolved, that the faculty go on record as firmly believing that in spite of uninformed and misguided criticism of this pilot program, the broad overall purposes merit and should receive full support from educators, from our community, and from the public at large."

Despite this praise, it would appear to the author of this History, in the bright light of hindsight, that the representatives of TALENT who made the arrangements with the school should have made clearer to them that this preliminary testing did not involve the Project TALENT Battery at all--or indeed any tests that would eventually be included in it.

---

* "Pseudo-battery" is the term I am using here to distinguish the set of instruments used from a formally organized battery available commercially or still in the developmental stage. The term was not used in 1958.
- Author

2. The Texas bonfire

Wayne Holtzman, who was a member of the Advisory Panel on Testing Problems, and was associated with the Hogg Foundation for Mental Health, in Houston, informed us that the Hogg Foundation was planning a county-wide testing program called the Houston Youth Study, which would include every 9th-grader not only in Houston but in many schools in the surrounding county, Harris County. Only noncognitive instruments (personality inventories, attitude scales, etc.) were being included. He offered to include in the battery any sets of noncognitive items (e.g. background information items, interest items) that we would like tried out. The offer was accepted, and history (the Baltimore brouhaha) almost repeated itself. Some of the Hogg Foundation instruments were personality inventories and attitude scales that were regarded by some of the local citizenry as undue invasion of privacy. There was at that time (and there may still be) an organization in Houston called "Minute Women." The Minute Women objected vociferously to the questionnaire items and demanded that all the answer sheets be burned. The Houston Board of Education, three members of which were reported to be active Minute Women, angrily denounced the study, and with only one dissenting vote, ordered that the answer sheets be destroyed "at once." Early the next morning, with reporters breathing down her neck, the Assistant Superintendent, who also was the chairman of the study committee, burned every answer sheet from Houston. An application by a concerned citizen for a temporary restraining order to prevent destruction of the data came several hours too late. The data were already ashes.

The episode resulted in headlines all over the country. It was generally referred to as "book-burning," and was widely deplored. In Houston, however, the press was largely hostile to the testing program. The *Houston Chronicle* editorialized: "It seems that no family skeleton is safe from the sociologists anymore."

Among the personality inventory items that caused resentment were the following:

- I enjoy soaking in the bathtub.
- A girl who gets into trouble on a date has no one to blame but herself.
- If you don't drink in our gang, they make you feel like a sissy.
- Sometimes I tell dirty jokes when I would rather not.
- Dad always seems too busy to pal around with me.

But the Youth Study had its defenders, too. The Child Welfare Section of the Community Council called on the Council's board to ask school board officials to cooperate with research organizations in the future in psychological and sociological studies involving field tests in the schools. The Community Council, however, included in its membership some school board members who managed to keep motions of that sort from passing.

Meanwhile the opposition to the Youth Study in Houston had stimulated similar opposition in surrounding suburbs where the Houston Youth Study battery had also been administered. However, there was no test-burning outside of Houston, and most or all of the county data reached the Hogg Foundation intact.

Again, however, as in the Baltimore Brouhaha, the instruments and items that were criticized were in no case Project TALENT's.

Chapter 3. ISSUES AND DECISIONS

This chapter is concerned with issues, mostly but not exclusively of a technical nature, on which decisions had to be made before the "main event", the 1960 data collection, could get under way--and with the decisions that were made concerning those issues. The issues fall into three categories: the sample, the design, and the battery.

A. The sample

As was indicated in Chapter 2, certain features of the sample were inherent in the basic idea of the project. The sample would have to be very large and it would have to be a probability sample. The probability sample feature meant that we could not settle for just a very large sample consisting of whoever wanted to participate (a variety of the "available cases" method); nor for a very large sample achieved by inviting schools selected in some nonrandom way (most likely because they were convenient) to participate, and cutting off the rolls whenever the desired number of acceptances had been received (another variety of the "available cases" method); nor for the "quota method". (The quota method is a modification of the available cases method whereby instead of just setting an overall sample size to be sought, the size of the sample is set separately for various categories of schools, but within those limitations the "available cases" approach is used.)

But the decision that the sample should be large and that it should be a true probability sample didn't answer all the questions. How large should the sample be? And just how should it be structured? How many grade levels should be included? Which ones?

Then, of course, there was the whole area of highly technical questions that needed answers. For instance, should the primary sampling unit be the student? the classroom? the school? the school system? the county? If the sampling unit is something larger than the individual student how many sampling stages should there be? If the sampling unit is the school, who should we cope with the problems of the junior high school? And in any event, should the sample be stratified? If so, on what? Should a uniform sampling ratio be used for all strata?

1. The Sampling Advisory Panel

The sheer magnitude and complexity of the sampling problem, combined with the fact that none of us, at that time, had had experience with a sampling problem of that magnitude, and none of us considered himself (or herself) an expert on the fine points of probability sampling, meant that we needed help from experts--a fact that was underlined when a member of the Cooperative Research Program's Research Advisory Committee raised a question on our sampling plans that seemed to require an answer from a sampling expert.

Because of that, and also because of all of our own unanswered questions, an advisory panel on sampling was established. Its four members are listed in Figure 3-1.

The panel met twice--first on 12 March 1959 and then on 26 September 1959, both times in AIR's Washington Office. The Office of Education was represented at the first meeting by Dr. J. William Asher (now on the faculty at Purdue, and a loyal user of TALENT data, via the Data Bank) and at the second meeting by Bill Asher again, and also by Howard Hjelm (who is still at the Office of Education, where he is now the Associate Commissioner for Adult Vocational, Technical, and Manpower Education). At the meetings, most of the issues mentioned above were considered, along with a few others.

FIGURE 3-1. Sampling Advisory Panel

Morris H. Hansen
    Assistant Commissioner for Research and Development
    Bureau of the Census

William G. Cochran
    Professor of Statistics
    Harvard University

Philip J. Rulon
    Professor of Education
    Harvard University

Frederick F. Stephan
    Professor of Social Statistics
    Princeton University

Considerable attention was devoted to the problem of sample size. Some members of the Cooperative Research Program's Research Advisory Panel had wondered whether we really needed as large a sample as had originally been planned (close to half a million students in over a thousand schools). Wouldn't a carefully chosen sample of, say, 50,000 serve the purpose at least as well? Some members of the sampling panel initially had the same doubt. But their doubts vanished when John Flanagan explained to them why the very large sample was central to his idea. He made it clear that for some of the major purposes of the project it would be necessary to analyze the data in terms of sample members' career fields (as identified in the follow-ups). For some career fields, such as physicist, anthropologist, meteorologist, the number of cases would be barely large enough to give useful results if we started with half a million students. The sampling panel members were won over to this viewpoint.

When the question of choice of sampling unit was tackled consensus was quickly reached in favor of our original plan--to select a sample of schools, and test all students in the selected grade range who were in the chosen schools. (In other words there would be no second-stage sampling of students or classrooms within school.) Admittedly, using students rather than schools as the primary sampling unit would have had some advantages from a theoretical viewpoint but it would have been utterly out of the question from any practical viewpoint.

As for the grade range to be sampled, it was quickly agreed that grades 9 through 12 would provide a useful range.

The next problem tackled was stratification. We (the project staff) knew it would be advantageous to stratify but we weren't sure what combination of stratification variables to use. There were a lot of possible stratification variables that we considered. Some had to be eliminated because it wouldn't be feasible to acquire the necessary information prior to sample selection; others would have to be eliminated just to cut down the number of stratification variables. Having too many would create practical problems as well as theoretical ones. (Classifying the universe of secondary schools into an enormous number of tiny cells would mean that there would be a lot of cells with so few schools in them that in the sample selection those cells would have to be very heavily oversampled to insure even one sample member.)

Among the stratification variables that we (the project staff) had been considering as possibilities were school control (3 categories--public, parochial, other private), region of the country, state, population of community, kind of community (urban, suburban, other small town, rural), socioeconomic level of the neighborhood served by the school, school size, grade range of the school, "school quality" (whatever that might mean), kind of high school (e.g., academic, comprehensive, commercial, vocational). Clearly, we weren't suffering from a paucity of ideas as to what to stratify on; the problem was an overabundance of ideas.

The Sampling Advisory Panel quickly confirmed that it would be highly desirable to use a stratification procedure in sampling; and they quickly endorsed our tentative plan to stratify on the public-parochial-private categorization.

As for geographical location they suggested geographical region, with the states arranged geographically within region, so that any "remnant" sets of schools (e.g., a remnant set consisting of 16 schools left over after one school had been selected at random from each set of 20) would "spill over" into an adjacent state.

A further refinement of the plan was introduced when a member of the panel pointed out that the very largest cities (cities with populations much in excess of a million) were more like each other than they were like the rest of their states; Chicago, for instance, resembled Philadelphia more than it resembled the rest of Illinois. That being the case, it was suggested that each of these largest cities be treated, for purposes of sampling, as if it were a state; and that together they should form a "quasi-region".

We pointed out to the panel that 5 percent of the schools had 50 percent of the students and 50 percent of the schools had 5 percent of the students. This led to the consideration of the idea of stratifying on school size. The idea was endorsed, and in connection with it, Dr. Hansen pointed out the importance of "oversampling" the very large schools (in other words, using a higher sampling ratio for such schools than for ones with fewer students). This would help reduce sampling error on student statistics (though not on school statistics). For similar reasons he recommended undersampling the very tiniest schools.

He also pointed out that to be able to get an empirical check on the magnitude of sampling errors it might be advantageous to select two parallel subsamples and compute statistics for them separately, before combining the subsamples.

The panel also agreed that it would be desirable, if possible, to stratify on some measure of "school quality" or of socioeconomic status (which were of course recognized not to be synonymous)--but nobody present ventured a suggestion as to what that measure might be.

One further topic discussed by the panel was what to do about schools that declined to participate. Should we select a replacement school? Dr. Hansen answered this question firmly in the negative. He pointed out that it would be every bit as satisfactory, from a theoretical viewpoint, and considerably less trouble from a practical viewpoint, just to increase the weight given to other schools in the same stratification cell to compensate, insofar as possible, for the missing school. He also stressed the fact that no participating school, whether it was a replacement school or just a more heavily weighted school from the originally selected sample, can wholly substitute for a school that refuses to participate, and that therefore it was imperative that we make every effort to get a near-perfect participation rate, for the initially selected sample.*

One important sampling problem, unfortunately, had been overlooked in preparing the agenda for the panel meeting. That was the problem of how to select the sample of junior high schools. I have no idea as to whether the panel would have come up with a good solution to this troublesome and nearly (though not quite) intractable problem if the question had been raised at the meeting. It wasn't, and so we shall never know. (To find out how we eventually handled the problem, see Chapter 7, section B.)

2. After the Sampling Advisory Panel meeting

The meeting of the Sampling Advisory Panel had been extremely helpful. It had confirmed the initial ideas on sample size and it had provided us with the outlines of a sampling plan, and with a strategy for coping with the problem of schools in the sample that might decline to participate in the project. But the details of the sampling plan remained to be worked out.

We were going to try to select our sample in such a way as to include about 400,000 to 500,000 students. The grades tested would be grades 9 through 12. The primary sampling unit would be the school—and there would be no secondary sampling. There would be three, or perhaps four, stratification variables:

---

* We took Hansen's advice to concentrate our efforts on getting the initially selected sample to cooperate; and those efforts proved very successful. (See Chapter 7, section B.)

1) Public-parochial-private
2) Geographical location
   (Region, and state within region, with the very largest cities treated as "quasi-states" combining to form a "quasi-region". A strategy for handling "remnant" groups of schools in a state was also suggested.)
3) School size
4) "School quality" or socioeconomic level of the school, or some combination of the two

There would be differential sampling ratios, primarily on the basis of school size, with the highest ratio being used for the largest schools.

But what cutting points would be used to define the school size strata? And what sampling ratios would be used? (Whatever ratios were selected, they would have to be ones that would yield the right sample size, in terms of both number of students and number of schools.) And what on earth were we going to use for a direct or surrogate "school quality" or SES measure that would be available before sample selection?

On the basis of a review of what kinds of information were available to us on the 90,000 IBM cards (3 per school) provided by the National Center of Educational Statistics (NCES) and containing data from their 1958-59 National Survey of Public Secondary Day Schools (see Chapter 7, section A), and some preliminary computer runs of a sample of these cards, John Flanagan made decisions on these basic unanswered questions. For the public schools four strata would be established on the basis of senior class size, with sampling ratios as shown in Table 3-1. For the nonpublic schools there would be stratification on only two variables: (1) Control (2 categories: "parochial"* and "other private"), and (2) location (region and state, including quasi-states and a quasi-region, as in the case of the public schools). From the preliminary computer runs it could be correctly inferred that the ratios shown in Table 3-1 would yield about the right number of schools and students.

TABLE 3-1. Sampling ratios and senior class size strata

School Control	No. of seniors	Sampling ratio
Public	400 or more	1:13
	100-399	1:20
	25-99	1:20
	Under 25	1:50
Parochial		1:20
Other private		1:20

After studying the types of data available on secondary schools, John (Flanagan) decided that the remaining stratification variable would be "retention ratio"--defined as the ratio of graduates to 10th-graders. This could be expected to be correlated with socio-economic status, and as a measure of the school's "holding power" it could be regarded as, in one sense, an indication of "school quality" (still undefined).

3. The 16-year-olds and the 15-year-olds

It will be recalled from Chapter 2 section B that funding had been secured from the Cooperative Research Program for supplementing the grade 9-12 probability sample with a probability sample of 16-year-olds not in grade 9, 10, 11, or 12. These were to be all the 16-year-olds in the school districts corresponding to about one-tenth (a representative tenth) of the public schools in the regular sample. These 16-year-olds, properly weighted and combined with the 16-year-olds in the grade 9-12 probability sample, would constitute a probability sample of the age group.

---

* The term "parochial" is used here as a kind of shorthand for what would more precisely be described as "parochial or diocesan".

The necessity for getting this representative tenth of the schools led to the decision by TALENT's Principal Investigator to adopt, in a considerably modified form, the suggestion Dr. Hansen had made at the sampling panel meeting, that we have two parallel samples. The modification consisted in splitting the schools in the sample into 10 subsamples ("subsamples 0-9") parallel in terms of the geographical stratification variable and the public-parochial-private stratification variable. One of these 10 subsamples (specifically "subsample 0") would define the school districts for the 16-year-old sample.

Shortly after this decision was made it was decided that for practical reasons it would be better to have a probability sample of 15-year-olds than of 16-year-olds. The principal reason for this change was that it became apparent that there would be too many 16-year-olds not in school to permit a good job of locating and testing them to be done within budget. Because of the compulsory education laws which in every state required school attendance until at least age 16, there would be far fewer 15-year-old dropouts. Thus a probability sample of 15-year-olds could be established more easily, and would serve the purpose at least as well as a probability sample of 16-year-olds would. Therefore permission was secured from the Office of Education to change the sampling plan in this one respect.

4. The 10 subsamples: an interim feature

The 10 subsamples had served one of their two functions when one of the subsamples was used to define the school districts for the sample of 15-year-olds not in grades 9-12. The second important function was to provide on a preliminary basis a sample one-tenth the size of the total sample, for use in analyses not requiring the full sample. But for that function the school subsamples were useful only at the beginning to make it possible to secure preliminary results, based on subsample 0, when only that subsample had been put on tape in a form suitable for data analysis. When all 10 school subsamples had been put on tape in master file form the need for the subsamples as such vanished, since we could get a much better 10% sample by selecting a representative tenth of students; this was easy to do on the basis of last digit of student number. Thus the ten school subsamples have hardly been used since good 10% student subsamples became available.

B. The design

Among the problems that were considered by the four advisory panels (see Figure 2-1) at one time or another were some that were concerned with the basic design of the study. For instance:

1. Should the testing be a one-shot event or should there be retesting in the same schools in subsequent years?
2. And if there is to be retesting, should it be of the same cohort at a higher grade level level? Or a different cohort, and the same grade level? Or both?

The recommendations of the Advisory Panels were uniformly in agreement with the original plans. There would be one-shot testing. Follow-ups would be by questionnaire only; there would be four--one, five, ten, and twenty years after the cohort's graduation from high school. There would be no routine retesting.*

---

* The special retesting described in Chapter 11 was not decided upon until the year it occurred (1963).

Obviously retesting would have had some advantages. But the decision against it was really the only possible one in view of the long-range purposes of the study. It will be recalled that the near-half-million sample that had been settled on was about the smallest that would give an adequate number of cases in some important but relatively infrequent occupations. But if there were to be any retesting, the per-capita costs would inevitably be increased substantially, and the sample size would have to be cut down accordingly.

C. To build or buy: The TALENT Battery

The principal problem to be solved in connection with the battery was where it should come from. Should it be a commercially distributed battery available *in toto* from some test publisher? Or should it be put together from a bunch of commercially distributed separate tests, available from one or more publishers? OR SHOULD A BRAND-NEW BATTERY BE DEVELOPED BY THE PROJECT STAFF, SPECIFICALLY FOR USE IN PROJECT TALENT?

The answer to these questions is telegraphed by the capitalization at the end of the preceding paragraph. The decision was--and had to be--that we would develop our own battery. That was really the only sensible decision--for a multiplicity of reasons, which are listed below:

1. In the first place it would avoid giving a tremendous commercial advantage to any one test publisher. Buying an intact battery from a single publisher would have given that publisher an enormous commercial advantage--not just in terms of the profits from selling half a million copies but also because of the increase in the value of the property as a result of all the normative and follow-up data that would be developed by Project TALENT, at no cost at all to the publisher. All the test publishers except the one whose battery was selected would have objected strenuously (and with good reason).

2. A specially built battery would insure that the tests would be new to all students alike. Thus distortion due to differential practice effects and to the fact that some examinees would already be quite familiar with a test while it would be brand new to others would be avoided. (A commercial test to which none of the sample had been exposed would not have been available.)

3. We needed a lot of very short tests rather than a smaller number of longer tests, since the battery would have to cover a very wide spectrum of abilities. Although with a battery of the sort we needed, some of the tests would be too short to give highly reliable results for individuals, a battery of this sort would still be better for research purposes than would a smaller number of longer more reliable tests. (Furthermore even though short individual tests may not yield highly reliable scores, they can be combined to produce very reliable composites.)

4. The final battery would have to be very tightly tailored not only physically but also in terms of test length, in order to make optimum use of the limited amount of testing time (2 days) that would be available in which to administer a very large number of instruments. The chances that each of a set of commercially distributed instruments would be precisely the optimum length would be virtually infinitesimal.

5. Unlike most users of commercial tests, we would not be dependent on the publisher for norms. (There goes one more argument for commercial tests, right out of the window!) With a probability sample consisting of nearly half a million students we felt confident that we could build our own norms, and that they would be better than the ones then generally available for commercial tests, both because they could be based on an enormous sample (nearly 100,000 cases per grade) and because it would be a genuine probability sample. It should be realized that in 1960 it was not really customary for test publishers to secure probability samples for norming. Samples were far more likely to consist of "available cases", and the number of

> available cases was more likely to be counted in thousands than in hundreds of thousands.

Even after the decision had been made that we would build our own battery, and thus not bestow an enormous advantage on a single test publisher, to the detriment of its competitors, some of the publishers were still a little uneasy about what effect Project TALENT might have have on their business. Reassurance was provided by a commitment on Project TALENT's part, that the TALENT battery would be used only for research--and that even then it would be made available only if the research study were such as to require the use of TALENT tests specifically--in other words if there were no tests commercially available that would do the job. The TALENT Battery would of course be made available at cost for use in equating commercial tests, so that they could benefit from TALENT's normative (and follow-up) data. These commitments allayed the uneasiness of those test publishers that had been uneasy, and we secured their support.

Chapter 4. NAMING THE BABY

Lots of people wonder where the name "Project TALENT" came from and what it means. They assume, because TALENT, used as the name of the project, is always spelled all in capitals, that it is an acronym. It isn't.

To begin with, the name of the project didn't come with the idea for the project. That idea goes a long way back, as explained in Chapter 2. The beginnings of John Flanagan's idea that AIR ought to do something along those lines took shape in 1956. But the developing idea was not dubbed Project TALENT until sometime in 1959. Originally it had a much more cumbersome name, "Research into the Identification, Development, and Utilization of Human Talents".

The Project Director, John Dailey, may have been the first person to realize how badly a short easily remembered name was needed to identify the project and make it stick in the public consciousness. After giving the matter some thought he announced one day that he had thought of the title: Project TALENT. He explained that the reason this name was appropriate was that everyone has some sort of talent--not necessarily one related to intellectual or artistic ability, but possibly a talent for assembly-line work or for tasks requiring physical strength or manual dexterity--or perhaps a talent for friendship. This broad (and humane) definition of "talent" makes the name of the project compatible with the character of the sample, representing virtually everybody, rather than a highly selected segment of the population.

The title has served its purpose well. The name Project TALENT rolls off many tongues all over the country. Inevitably, however, there has been some confusion among people who have heard the project name but know nothing whatever about the project. It has been confused occasionally with enterprises designed to identify high school graduates in the top levels of intellectual ability--for instance the Westinghouse Science Talent Search, or the National Merit Scholarship competition. But this kind of confusion is of little consequence, since it is so easily cleared up.

Somewhat more prevalent than the misconception about the implication of the word "talent" as used in the project name is the misconception mentioned earlier, that "TALENT" is an acronym. And numerous attempts have been made, mostly by uninvolved bystanders, to develop an *a posteriori* acronym. The author of this History offers the following demonstration that such efforts are to no avail:

T	Totally
A	Acronymless
L	Large-Scale
E	Educational
N	Nationwide
T	Testing

Chapter 5. BUILDING THE BATTERY

A. Schedule

The schedule for getting from the preliminary phase to a published battery in final form was going to be an extraordinarily tight one. That much became apparent as the delays in securing funding continued. It was clear that if the goal of doing the testing phase in the spring of 1960 were to be met, test development would have to be begun before funding was received, and so that is what was done. It was a gamble, but it was one that paid off; it made it possible to meet the schedule for a spring 1960 testing. (It was felt that 1960 was a particularly advantageous base year because it would coincide with the Decennial Census, thus facilitating all sorts of comparisons.)

B. "State-of-the-art"

It had been decided early that Project TALENT would develop a "state-of-the-art" battery. This meant that it was to consist primarily of well-established types of tests and inventories rather than novel or experimental ones about which little was known. But the instruments we were to build should of course be as good specimens of their kind as could possibly be achieved. (If we were talking about athletes, "state-of-the-art" would mean the Olympic champion, not the typical Sunday athlete.) That was the rather lofty goal. The degree to which it was achieved--indeed, whether it was achieved at least partially--cannot be asserted with any confidence. A man's reach--a person's reach--a project's reach--should exceed his (his or her) (its) grasp.

The foregoing discussion of goals in developing the TALENT battery is motivated by the fact that the term "state-of-the-art" is misunderstood by many people, who confuse "state-of-the-art" with "mediocre," rather than understanding it to mean the most advanced state-of-the-art; and whether achieved or not, that was what we were aiming at!

C. Content and organization of the battery

Decisions on content and organization of the battery were based in part on the test materials assembled by Bob Craig during his year of planning, but mostly on the recommendations of the test panel and on the judgment of the staff members especially knowledgable in the area. Chief among these, of course, was the Principal Investigator, John Flanagan, whose expertise had been thoroughly demonstrated by, among other things, his former roles as Associate Director of the Cooperative Test Service and as Colonel Flanagan, Director of the Army Air Force Aviation Psychology Program during World War II.

It was established early in the planning of the test battery that it would include, among other things, a long, multi-purpose Information Test, achievement tests in math and English, a reading comprehension test, and tests of various other well-established aptitudes and abilities.

Permission was received from John B. Carroll, the principal author of the Psi-Lambda Language Aptitude Test (now called the Modern Language Aptitude Test), to develop parallel forms to the three paper-and-pencil subtests of the Psi-Lambda. (The other two of the five subtests could not be handled within the Project TALENT framework because they required use of a tape recording.) The decision to use three tests of the Psi-Lambda type was made on the recommendation of Marion Shaycoft, who had observed use of that instrument in an entirely different context and had been much impressed by the content validity of the test, its demonstrated effectiveness, and the novelty and ingenuity of the approaches.

Speaking of ingenuity, John Flanagan's FACT Battery (Flanagan Aptitude Classification Tests) included a promising Ingenuity test. It was decided to handle this test on the same basis as the three Psi-Lambda subtests--in other words to construct a parallel form for inclusion in the TALENT battery.

In connection with the parallel forms of these tests from other batteries, certain name changes were decided upon. The FACT Ingenuity

Test became the TALENT Creativity Test. The Psi-Lambda subtests named "Paired Associates," "Spelling Clues," and "Words in Sentences" became "Memory for Words," Disguised Words," and "Word Functions in Sentences," respectively.

There were at least three departures from the decision to use only tests of well-established types rather than trying out experimental instruments. Henry Chauncey, the president of Educational Testing Service, who was a member of our Advisory Panel on Testing Problems, recommended that we include three tests that ETS had been using on an experimental basis and considered promising. One was called "Social Judgments" and required the examinee to answer a series of two-choice questions where the task was to indicate by choosing between two equally favorable (or equally unfavorable) adjectives the type of person he liked better (or disliked less). The score, which represented speed of making one kind of decision, was simply the number of items answered in a very limited time. (Of course, the nature of the scoring procedure is not revealed to the examinee.) The second test, which was called "Sentence Completion", called for studying a series of short sentences of a rather bland or neutral nature for a specified period to memorize them verbatim, and then filling in blanks with the missing words. The third test was called "Following Directions". We decided to try these tests since they sounded as if they might be useful and since they required relatively little time and therefore could be squeezed into the battery without much trouble. Our version of the "Social Judgments" test is called "Preferences;" our version of the "Sentence Completion" test is called "Memory for Sentences;" we had to make major revisions in the item format to change the memory items from write-in to multiple choice, but we believe the new format that we designed retained the essential features of the ETS test; i.e., it measured deliberate memorization rather than incidental memory, underlearned rather than overlearned material, and, most important of all, recall rather than recognition (despite the multiple-choice format)!

The "Following Directions" test which, like the Sentence Completion test, required a new item format to fit our requirements that every test item be machine-scorable, did not survive the item analysis tryout. (The other two tests did survive, and are in the final battery.)

As a matter of fact, in addition to the three "experimental" tests developed and tried out at Henry Chauncey's urging there was at least one additional test of a type that might, by some definitions, have been considered "experimental" rather than "state-of-the-art." The author of this History, however, who was also the author of the test in question, contends that the test in question, which happens to be the Arithmetic Reasoning Test (Part I of the 3-part Mathematics Test), is, if not a state-of-the-art test, demonstrably better than such a test. To be a little more specific, the Arithmetic Reasoning Test written for the TALENT Battery is virtually computation-free. Most conventional arithmetic reasoning tests at that time required (and most of the ones now being developed* still require) that the examinee not only know the correct method of solving "word problems," but demonstrate that knowledge by actually doing the computation necessary to find the answer. Thus a student good at arithmetic reasoning but poor at computation might get an unduly low score which would (unfairly) cast doubt on his or her reasoning ability. The TALENT Battery Arithmetic Reasoning Test clearly separates reasoning from computation, and requires only the former. Thus if it was not precisely a state-of-the-art test, perhaps it advanced the state of the art (at least a little bit).

Prior to developing the tryout forms of the tests a very complete rationale was written for each test. The rationale explained what was to be measured and how it was to be measured, provided detailed specifications for the test items, indicated the distribution of the levels of item difficulty to be sought, and presented sample items. The concept of the rationale, developed largely by John Flanagan, should, in AIR's collective thinking, be an integral part of any test development activity.

---

* The PCG Quantitative Reasoning Test is an exception. See Chapter 23.

D. The battery-building crew

Building the experimental form of the battery was an in-house activity, carried out largely by project staff, rather than a set of tasks to be farmed out to consultants, subcontractors, and other external personnel or agencies.

John Flanagan played a central role--the central role--in planning the outlines of the battery's content and organization. He worked closely with the staff members responsible for the various parts.

1. The test builders

Marion Shaycoft was the editor-in-chief of the aptitude and achievement tests. In this capacity, and in close conjunction with John Flanagan, she played a major role in establishing the outlines of the entire battery and making it fit within the physical and temporal constraints (test booklets and answer sheets representing the former kind of constraint and the two-day time limit for administering the battery the latter). She planned the details of all the tests, wrote the rationales for them, supervised the item writing, and edited the results. (John Flanagan, of course, did the final review of all rationales and all tests.) As things turned out, because the test development staff was quite small Marion Shaycoft also wrote many of the tests almost in toto and wrote large parts of most of the others. She was fortunate, however, that though members of the project staff available for this phase of the work were not numerous they did include at least one staff member who was both experienced and very competent in test development. That was Jo Edwards. She was the associate editor and a mainstay of the test development staff. Among other things, she is wholly responsible for the item format and all the items in the main section of the Punctuation Test.) It was decided to farm out those few segments of the test-writing that we felt could best be handled on that basis. The item-writing handled that way included a few of the areas in the Information Test--for instance the items on auto mechanics, an area about which neither Jo Edwards nor Marion Shaycoft nor any other staff member on the premises felt at all knowledgeable--and the mechanical reasoning items--which were drafted by Mary Willis, a highly experienced and expert test writer who at that time was not on the AIR staff but was later to join it. (Mary had been a test editor many years before, at the Cooperative Test Service (CTS)--one of the predecessors of the Educational Testing Service--when John Flanagan had been the Associate Director. It was at about the same time and in the same organization (CTS) that Marion Shaycoft had been initiated into the fine art of test writing under John Flanagan's aegis. Geraldine (Gerry) Spaulding, who was to be one of the reviewers of our tests, and a consultant on our questionnaires, was a test editor in the same organization, CTS, at the same time.

John Flanagan assumed personal responsibility for writing the Creativity Test, since it was to be modeled on the Ingenuity Test in his FACT Battery. He also wrote quite a few of the sports information items, as did John Dailey, since neither Marion Shaycoft nor Jo Edwards felt confident of her ability to deal with the technical aspects of such activities as football. John Dailey took personal responsibility for writing the Information Test items in the areas of Hunting and Fishing. (It is perhaps worth mentioning at this point that the Hunting Information score turned out to have a larger proportion of unique reliable variance than any other test in the battery. And for those readers who may be a little fuzzy as to whether unique variance is a good thing or a bad thing, perhaps we should clarify matters by stating flatly that unique variance is a very good thing indeed, provided it is also reliable variance, as it was in this instance.)

Probably because she had been instrumental in getting Jack Carroll's permission to prepare tests parallel to three of his Psi-Lambda subtests, Marion Shaycoft wound up writing the three tests. In granting permission Jack had imposed the entirely reasonable condition that he be allowed to review the tryout forms of the test and that his approval had to be secured before use. This proved to be an advantageous arrangement, since he readily gave his approval of the three tests, thus giving us the assurance we wanted that our tests were indeed parallel to his. This was particularly reassuring in the case of the Word Functions in Sentences Test, since its nature was such that it needed to have somewhat vague directions. For this test, whether one could comprehend the instructions was a legitimate component of the test score. (This is not true on most tests in the cognitive domain. On most such tests the instructions are supposed to be crystal-clear, so that all examinees can be assumed to understand them, thus permitting any reduction in score to be attributed to inability to cope with the items themselves, not with the directions.)

It was indicated at the beginning of this section that the test construction was largely an in-house activity. One partial exception to this should be mentioned. Because of our extremely tight schedule and because the art work on items requiring diagrams (visualization items, abstract reasoning items, object inspection items, mechanical reasoning items) is very time-consuming, it occurred to us that it would be worthwhile to explore whether the U.S. Civil Service Commission had available any good items of these types which they could spare for us. Accordingly Jo Edwards and I conferred with Civil Service Commission personnel (Albert Maslow, Harold McAdoo), who proved most cooperative. After looking into the situation they reported that they had quite a few visualization items of the "paper-folding type" (in Project TALENT's terminology Visualization in Three Dimensions items) which had been item analyzed and were known to work well, which they could let us have. (They found they did not have suitable items for any of the other tests.) We accepted their offer and included a substantial number of their "paper folding" items in our experimental forms. Accordingly some of the items in the final form of the TALENT Visualization in Three Dimensions Test are in there through the courtesy of the Civil Service Commission, while others were developed by project staff.

2. Questionnaires for students

Dave Orr was assigned responsibility for developing and editing the SAI (the "Student Activities Inventory"--this was a code name for "personality inventory").

Izzy Goldberg had a corresponding role in regard to the other two student inventories--the SIB ("Student Information Blank," which provided background information) and the Interest Inventory.

3. School questionnaires

Dave Orr was responsible for developing and editing the General School Characteristics Questionnaire. Izzy Goldberg had the same role with respect to the other two school questionnaires--the Guidance Program Questionnaire and the Counselor's Questionnaire. One other person played a major role, on a day-to-day basis, in the development of these school-level questionnaires. That person was Dr. David F. Votaw, Sr. Dr. Votaw, who had been John Dailey's professor many years before, at the college from which he received his bachelor's degree, Southwest Texas State Teacher's College, had been invited to spend several months in Washington during the 1958-1959 period as a consultant to the fledgling project. In this capacity Dr. Votaw concentrated his efforts on the school questionnaires and made many important decisions in regard to them.

E. Tryout and item analysis

The experimental tryout was carried out in the spring of 1959, in 11 high schools varying widely in size, kind of community, and section of the country. All students in grades 9-11 were tested (although each student took only a part of the battery). Altogether, almost 6,000 students were involved; each form of each test was tried out on about 1,000 to 1,600 students.

Partly on the basis of the item analysis but chiefly because limitations on testing time required it, four tests were eliminated from the battery. The rest were revised and submitted to members of the Advisory Panel on Test Problems (see Figure 2-3 for their identities), as well as to other external reviewers; specifically, Paul Diederich, Gerry Spaulding, Mary Willis, Machlin Thomas, Max Engelhart, Paul Dressel. Their reviews were almost uniformly very complimentary. It will be recalled, too, that Jack Carroll reviewed the three tests based on his Psi-Lambda Battery. Thus all our tests received a very thorough review by nationally recognized experts, and passed with flying colors!

F. The end of the process

After the final review, final editing, and final revision of the final forms, we were ready for the penultimate steps--putting all the tests and other materials together and developing essential auxiliary materials such as a manual of instructions for test administration, and in general forming a cohesive battery out of a collection of separate components. The battery that resulted was organized in five separate test booklets: Booklets A, B, C1-x (an "expendable" booklet, i.e., one on which the students were permitted to write), C1, and C2. There were also five answer sheets: A, B1, B2, C, and Record form Z (the form containing basic identifying information such as name and address).* Answer sheets corresponded to the test booklets which had the same

---

* Record Form Z is shown in Figures 5-1a (front) and 5-1b (back). We hoped to get a Record Form Z for everybody in the class--even absentees--in order to complete our probability sample. Therefore the instructions in Figure 5-2 were incorporated in the Local Coordinators' Manual.

*TALENT TESTING PROGRAM

Record Form Z

* removed due to copyright restrictions

Figure 5-2. EXTRACT FROM LOCAL COORDINATORS' MANUAL

SPECIAL TREATMENT OF STUDENTS WHO DO NOT TAKE THE TESTS

1. Teachers should be instructed to fill out a Record Form Z for each student who is not tested, and to assign (but not hand out) an identification card to each such student, so that he will be identifiable by a six-digit "testing number."

2. On Record Form Z, Side 1, just to the left of the heading "TALENT TESTING PROGRAM" is a column of 39 answer space circles with the injunction "DO NOT MARK IN THESE SPACES."

   This column is to be used by the teacher or Local Coordinator (or in some cases by the Regional Coordinator) to indicate the reason the student was not tested. One circle should be blackened in this column, using the following code:

   Codes A - E: Absence

   A. Excused absence: illness

   B. Excused absence: other

   C. Absence: habitual truant

   D. Absence: reason unknown; has been to school at least once in the past month

   E. Absence: reason unknown; has not been to school at all in the past month

   Codes J - Q: Excused from taking tests (not necessarily absent)*:

   J. Foreign student: non-English-speaking

   K. Completely illiterate (though probably not mentally retarded)

   L. Completely illiterate (presumably because of severe mental retardation)

   M. Blind

   N. Nearly blind

   O. Paralysis, palsy, amputation, or some other severe handicap that prevents writing

   P. Some other severe physical handicap that prevents taking the tests

   Q. Temporary physical disability (e.g. broken arm)

   Codes S - W: These codes apply only to the "special (15-year-old) sample," and primarily to those cases in it who are in college or not enrolled in any school. In general they do not apply to 15-year-olds in elementary school, although they may in some cases. These codes (S - W) should not be used if any of the codes J - Q is applicable.

   S. Refused to take tests

   T. Agreed to take tests but then failed to appear at scheduled time

   U. In reform school, institution for delinquents, jail, etc. Permission for testing was withheld by the custodial institution.

   V. Parent or guardian refused to permit the 15-year-old to be tested

   W. Some other reason

3. In general, codes A - Q will be used by the teacher or Local Coordinator, to indicate why students enrolled in the schools were not tested. Codes J - W will be used by the Regional Coordinator to indicate why "special sample" cases were not tested.

---

*Before assigning any of these codes, the teacher or Local Coordinator should be thoroughly familiar with the contents of Paragraph IV on page 3 of the Teachers' Guide. In this paragraph principles to follow in determining whether to excuse a student from the tests are discussed.

letter designations--so it is clear that even though there were five of each there wasn't a one-to-one correspondence. In fact the whole thing was like a giant jig-saw puzzle; getting it all to fit together in a technically sound physical arrangement (e.g., with the test booklets organized in such a way that test items were not visible when test instructions were being read) was the very complex penultimate step in producing the battery. Since the volume involved was so large that one extra page in a test booklet could increase costs tens of thousands of dollars, we also had to be economical of space. The ultimate step was the printing--an enormous task in itself. The answer sheets were printed by MRC, of course, since only MRC was set up to produce answer sheets that would fit their machines within the very small tolerances allowed. Everything else was printed by the lowest bidder equipped to do a competent job--the Cuneo Press, in Philadelphia. It was done on such a tight time frame that it was necessary for someone (me) to stay on site in Philadelphia a couple of weeks to answer questions, look at galleys as soon as they were ready, and in general function as an interface* between TALENT headquarters in Washington and the massive printing operation. Eventually it got done--and we wound up with what we were told amounted to 18 freight carloads of printed paper! (Arrangements had been made for the materials to be shipped direct from the press to the schools all over the country, thus bypassing our headquarters. There was no way 18 freight carloads of paper could have been squeezed into our very tiny offices in Washington.)

G. What we ended up with

Readers who are interested in getting some idea of what sorts of instruments constituted the final battery and what sorts of measures they yielded are invited to look at Appendix A. Appendix A-1 shows, in tabular form, details concerning the final battery. Appendix A-2 consists of thumbnail descriptions of the principal tests.

---

* But in 1960 I didn't know I was an "interface"! The word interface used in that sense had not yet entered the general vocabulary.

- Author

Chapter 6. PRE-"CRITERION-REFERENCED" CRITERION-REFERENCED TESTS

(The TALENT "Domain Tests")

A. The concept

The term "criterion-referenced tests" was coined by Robert Glaser in 1963 to contrast a certain kind of test score interpretation with a "norm-referenced" interpretation.

The basic Project TALENT battery, the construction of which was discussed in the previous chapter, consisted, of course, of conventional norm-referenced tests. None of the tests was, technically speaking, a criterion-referenced test (although a criterion-referenced interpretation possibly could be imposed on some of them, on an *a posteriori* basis). A criterion-referenced test is a test in which the scores have some sort of absolute meaning, rather than being meaningless except insofar as they can be related to the performance of some "norms group." Criterion-referenced scores (scores with absolute meaning) are numbers to which an interpretation such as one of the following can be attached:

> A score of 96 means that the examinee can correctly compute 96 percent of all products of a 2-digit number by a 1-digit number.
>
> A score of 60 means that the examinee can read the lead article in the Washington Post with "60 percent comprehension."
>
> A passing score on the road test means that the driver can satisfactorily perform all the maneuvers one needs in one's repertoire in order to be a safe and competent driver.
>
> A score of 120 on this vocabulary test means that one has an active vocabulary of 60,000 words.
>
> A passing score on this end-of-unit examination means that the student has mastered the content of the instructional unit well enough to be able to cope with the next instructional unit, in which he will have to utilize the information he is supposed to have already acquired.

The above examples of criterion-referenced interpretations of test scores contain no particularly novel concepts--no concepts of sorts that were unheard of before the term "criterion-referenced" was coined in 1963. Clearly the notion of criterion-referenced tests has been around a lot longer than the name for them.

As a matter of fact, after construction of the regular Project TALENT Battery was well under way, John Flanagan proposed that in addition we develop criterion-referenced tests in several areas and equate them to the corresponding tests of the regular TALENT Battery, so that criterion-referenced performance (on the new tests) could be given norm-referenced interpretation (via the regular TALENT tests). Of course in presenting this idea John did not use the term "criterion-referenced". Remember, this was 1960, and the term did not exist yet. He called the proposed new tests "domain tests".

B. The tests

It was quickly agreed that the areas in which we should develop "domain tests" were Reading Comprehension, Vocabulary, and Spelling. There were to be two reading tests--one to measure comprehension of works of fiction and the other to measure comprehension of magazines. The vocabulary test was to be a measure of size of vocabulary and the spelling test a measure of the proportion of words one could spell correctly, in some suitable carefully defined population of words.

John Flanagan assumed personal responsibility for developing detailed plans for the two reading tests and getting them written, and Marion Shaycoft thus found herself with similar responsibilities in relation to the other two domain tests (Vocabulary and Spelling).

John arranged with Mary Willis (the same Mary Willis who, in the previous chapter, was identified as the author of our Mechanical Reasoning items and who also happened to be an experienced and expert writer of reading comprehension tests) that she would write the two domain Reading Comprehension Tests. For Test A-1 (Authors), John and Mary jointly picked out 10 eminent writers of fiction, ranging from authors whose writings were very easy to understand to authors generally considered difficult. Specifically, the authors were Louisa May Alcott, Robert Louis Stevenson, Rudyard Kipling, Willa Cather, Sinclair Lewis, Jules Verne, Jane Austen, Joseph Conrad, Feodor Dostoevski, and Thomas Mann. (I leave it to the reader to decide whether that list is supposed to be arranged from easy to hard or from hard to easy.) Similarly for Reading Comprehension Test A-2 John and Mary picked 10 magazines, ranging from movie magazines to prestigious slick-paper publications. (Actually 11 magazines were picked, but two of them, which were both movie magazines, were treated as a single publication.)

For each of the 10 authors two books were selected, and one passage per book was chosen for the test. Similarly two passages were chosen from each of the 10 magazines. Five multiple-choice test items were written per passage, in accordance with a specific pre-defined pattern; for each passage there was to be one item of each of the following types:

1. An item testing _general_ _understanding_ of the passage--comprehension of its general idea or intent;
2. An item testing comprehension of one of the _simplest_ points in the passage;
3. An item testing comprehension of a point of _average_ difficulty (relative to the difficulty of the passage);
4. An item testing comprehension of the most _difficult_ point in the passage;
5. An item testing _appreciation_ of the passage, or an _application_ of the ideas in it, or an _inference_ drawn from it.

As for the Vocabulary Domain Test, which contained about 300 items, it was designed to provide an estimate of the number of _word-meanings_ (not words) in the examinee's "comprehension vocabulary" (the meanings he _understands_, as opposed to his "active vocabulary," the meanings he _uses_ in his speech or writing).

The 150-item Spelling Domain Test provided an estimate of the number of words, among the most frequently used 5,000, that the examinee could spell correctly.

All these tests required very detailed rationales, with precise rules for sampling (of passages in the case of the Reading Tests, word meanings in an unabridged dictionary in the case of the Vocabulary test and words in the Teacher's Word Book of 30,000 Words*, in the case of the Spelling test), and with very specific rules for item construction. Results and further details on the tests' construction are reported in the following publications:

For the two reading tests:

Flanagan, J. C. Project TALENT: Preliminary findings. Paper presented at a joint meeting of the American Educational Research Association and the American Association of School Administrators, Atlantic City, February 20, 1962.

Flanagan, J. C. Implications of Project TALENT for research in the teaching of English. In Steinberg, E. (Ed.), Needed research in the teaching of English. Proceedings of a conference at Carnegie Institute of Technology, May 5-7, 1962.

Flanagan, J. C. Maximizing human talents. The Journal of Teacher Education, June 1962, 13(2), 204-215.

Shaycoft, Marion F. Chapter 3 in Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycoft, Marion F., Orr, D. B., Goldberg, I., and Neyman, C. A., Jr. The American high school student. (Final report to the U.S. Office of Education, Cooperative Research Project No. 635.) Washington, D. C.: University of Pittsburgh, Project TALENT Office, 1964.

For the spelling test:

Shaycoft, Marion F. Chapter 3 in Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycoft, Marion F., Orr, D. B., Goldberg, I., and Neyman, C. A., Jr. The American high school student. (Final report to the U.S. Office of Education, Cooperative Research Project No. 635.) Washington, D. C.: University of Pittsburgh, Project TALENT Office, 1964.

---

* Thorndike, E. L., and Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

For the vocabulary test:

> Shaycoft, Marion F. Development and analysis of a test to estimate size of vocabulary. Paper presented at the American Psychological Association convention, Chicago, September 7, 1965; revised February 26, 1968. Palo Alto, California: American Institutes for Research, 1968.

Chapter 7. EXECUTION OF THE PLANS

This chapter tells what was actually done--beyond putting the finishing touches on the Project TALENT Battery (the development of which is described in Chapter 5)--in the fall of 1959 and the spring of 1960 to convert Project TALENT from a concept to a reality.

A. Selecting the sample

Late in the fall of 1959 we were ready to select the sample. The sampling procedure had been decided upon (see Chapter 3, Section A). The National Center for Educational Statistics (NCES), at that time a branch of the Office of Education, had made available to us preliminary decks of cards containing the basic information we would need in our stratification procedures (and a lot of other information) from their 1958-59 National Survey of Public Secondary Day Schools. This was eventually to become a directory, providing basic information about all public secondary schools in the United States (both junior high schools and senior high schools). There were, in very round numbers, some 30,000 schools in this population, and NCES had six IBM cards chock-full of information about each school. We needed three of the six cards, so NCES obligingly made available to us over 90,000 IBM cards. If the same thing were being done today, all that information would probable be on just one 11½-inch reel of tape--but in 1959 cards were the basic unit, and one never bypassed them by going directly from data to tape. Anyhow perhaps the 90,000 cards were good practice for the time when about five million cards containing raw data from the administration of the TALENT battery (see Section E of this chapter) would be deposited in the lap of our programmer at the University of Pittsburgh Computation and Data Processing Center. But that is another (and longer) story.

To get back to the sampling, our staff programmer, Gary Lotto, wrote a program to select the sample on the IBM-650 (a terribly primitive computer by contemporary standards, but regarded in 1959 as practically the last word in computer technology).

He ran the 90,000 cards and lo and behold, his program produced a printout listing about _____ schools--the initial _____ public school primary sampling units. These were of course all senior high schools--i.e., schools that went at least as far as grade 11*--since that was how we had defined the "primary sampling units". The corresponding junior high schools (the "feeder schools") were to be identified later, after the cooperation of the senior high schools had been secured.

fill in

To select the sample of parochial schools, and of other nonpublic schools, we secured lists of these two populations of schools. The National Catholic Education Association provided the list of parochial schools and NCES gave us a list of the other nonpublic schools.

Since neither the parochial schools nor the other nonpublic schools were on IBM cards, selection of those parts of the sample was done clerically rather than by computer. The procedure, however, was identical, except, of course, that by prior decision the only stratification variable used for these two categories of school was geographical location. Randomization was effected through the use of tables of random numbers. In selecting these nonpublic schools, in effect humans were imitating a computer (rather than vice versa).

As for the population of public secondary schools, it turned out that even the 30,000-odd schools represented by the 90,000 cards provided by NCES did not constitute the complete population. Since NCES's 1958-59 survey was still in progress, questionnaires completed by schools were dribbling in to them. These late arrivals were of course not yet on IBM cards, but they could not be ignored. Therefore whenever NCES had accumulated a substantial number of these additional schools they notified us by phone. A project staff member would then go down there** and, again "imitating a computer", would select a sample. The very last dribble consisted of three "late late schools", for which questionnaires turned up at the very last minute--as a matter of fact not long before the period set aside for conducting the testing was almost over. Using a table of random numbers we selected one of the three, and fortunately, despite the short notice it agreed to participate. For students in this single late late addition to the sample a weight of 3.0 (the reciprocal of the effective sampling ratio) had to be assigned, rather than the more usual 13.0, or 20.0, or 50.0 (or a still higher number, to compensate for nonparticipating schools).* Users of the TALENT tapes encountering for the first time the record of a student or school with a weight as low as 3.0 almost invariably are convinced it is an error in the tape. It isn't.

---

* Grade 11 was used in the definition, instead of grade 12, to allow schools from 11-year school systems, if any, to be included. (These 11-year systems, which had been prevalent in the South, were rapidly being phased out at the time, and none was selected.)

** Fortunately, Project headquarters were in Washington at the time, so it was easy to get to NCES.

A supplementary source of public schools was the Internal Revenue Service. The IRS had a list which included some schools for which NCES did not have questionnaires. Accordingly, we sampled from the IRS list too.**

But in spite of the physical problem of juggling 90,000 cards, and the nuisance of coping manually with late additions and with nonpublic schools, that was really the easiest part. Now we came to the hard part--or at least what we anticipated would be the hard part: getting the cooperation of the selected schools (and their feeder junior high schools). The next section of this chapter tells, among other things, about the pleasant surprise that was in store for us.

B. Getting the cooperation of the selected schools

As things turned out, getting cooperation was a lot easier than we had anticipated (and probably a lot easier than it would be today). This was in part because of the careful groundwork that had been laid in the years from 1957 to 1960 (see Chapter 2) and partly because of the dedication and competence of our Regional Coordinators.

---

* See Chapter 15 for a discussion of weights.

** Further details about the sampling procedure can be found in Chapter 3 of "Project TALENT: Designing the Study", by Flanagan, Dailey, Shaycoft, Gorham, Orr, and Goldberg (1960); or in Chapter 3 of "Design for a Study of American Youth" by the same authors (1962).

We had appointed about 90 Regional Coordinators--mostly members of Psychology or Educational Research faculties of colleges and universities all over the country. (The Regional Coordinators are listed in Chapter 8, Section C.) It was these Regional Coordinators who contacted the school superintendents and principals, told them about Project TALENT, and made them eager to participate. The activities of the Regional Coordinators were coordinated in project headquarters by Bill Gorham, who was in charge of getting the selected schools to cooperate and getting the tests administered.

First the good news!

Although the task of getting the schools' cooperation was not without problems, the overall success rate was an astonishing 93 percent. And in 19 states, including all six of the New England states, it was actually 100 percent!

As a matter of fact we got more than we bargained for in the way of cooperation. About one in 13 of New York City's high schools and about one in 20 of Chicago's had been selected. But headquarters of the New York City school system included a Bureau of Educational Research, headed by the very knowledgeable J. Wayne Wrightstone. Wrightstone and his staff pointed out that it would be very difficult, perhaps impossible, to disrupt an entire high school as large as those in New York City, for two whole days. Fortunately Wrightstone recognized the value of Project TALENT and wanted to participate. He immediately recognized what we had known right along--that if one in 13 of New York's schools was a good sample (as indeed it was) one in 13 of the students would be a superb sample. The reader will recall, from the discussion in Chapter 3, Section A, that we had decided to make the school, rather than the student, the primary sampling unit, not because we felt that that would have any theoretical advantages (which of course it wouldn't) but for purely practical reasons.

Thus we were absolutely delighted when Dr. Wrightstone proposed testing every 13th student in every secondary school (both junior and senior highs). And we were even more pleased when, somewhere along the line, the plan changed from every 13th student to every 12th student.

Authorities in the Chicago school system decided to follow suit, on a somewhat more restricted basis. Instead of the two schools that had originally been selected from the Second City's 38, the Chicago school authorities proposed that we select 10 times as many (i.e., 20 schools). They would then administer the TALENT Battery to every tenth student in those schools. Again, we were delighted, and again we acceded rapidly to this suggestion.

(It should be noted that via appropriate modifications in weights the additional schools tested in New York and Chicago could be treated as additional primary sampling units.)

Despite the better-than-we-asked-for cooperation from the two largest cities, little* Knoxville, Tennessee and the surrounding county put them to shame, by agreeing to test every student, in a five-grade range (i.e., grades 8-12 rather than the usual 9-12 range) in every school--public, and parochial, and private--and not just in the city of Knoxville but in the entire county. This coup was achieved primarily through the energy, initiative, and foresight of our Regional Coordinator in the Knoxville area, Louise Cureton. It was she who realized what a marvelous resource this would be for future research, and it was she who took the lead (after clearing the idea with Project headquarters) in selling the Knoxville and Knox County school authorities on the idea. Thus for that area we had the entire grade 9-12 population (plus grade 8) rather than just the two public schools initially selected.

---

* Little, at least, in comparison with those two metropolitan giants, New York and Chicago.

In addition to the more-than-we-expected instances of cooperation described above, there were other cases as well where school authorities asked whether a school that had not been selected for the sample might still be included. In many cases this occurred where there were only two schools in a community and if only one of them had been selected (as was likely to be the case) the school authorities felt it might be awkward to test just that one and not both. This was particularly likely to be the case in southern communities, where there might be one white and one black school (in 1960 there was still quite a bit of *de jure* segregation in the South despite the fact that it had been declared unconstitutional in 1954). Whenever a school system specified, as a condition of their permitting the selected school to participate, that we would have to allow the other school to participate too, we agreed. But since in most cases including the extra schools in the sample could have destroyed the sample's probability-sample characteristics, which would have been disastrous, we called the extra schools "non-probability schools". This meant that data from those schools would be left out of any data analyses that called for a probability sample. The non-probability data could, however, legitimately be included in any analyses that did not require the probability sample feature.

Actually most of the non-probability schools might have legitimately been included in the probability sample if a procedure developed by the author of this "History" had been applied. But that procedure, which is described in Appendix D1, wasn't even developed until about 15 years later.

Other non-probability cases occurred in some schools in the probability sample, where the school principal wanted to test eighth-graders, as well as the usual four grades, 9-12. When this request was made, we again acceded, and merely designated the eighth-graders "non-probability cases".

In addition to the "self-invited" nonprobability cases, the reader will recall that we had decided that in one-tenth of the school districts corresponding to probability-sample public schools we would test all 15-year-olds not in the grade 9-12 range. Many of these turned out to be in grade 8; some were in grade 7; and a few were in even lower grades. Some, of course, were not in school at all, but these were particularly hard to locate, since essentially they were truants, all states at that time requiring school attendance until at least age 16. In any event these 15-year-olds not in the grade 9-12 range clearly were not in the grade 9-12 probability sample--although, equally clearly, they were in the probability sample of 15-year-olds. In schools in which there were a very large number of 15-year-olds to be tested in grade 8, school authorities quite often opted to test the entire grade. That gave us still more eighth-graders in neither the grade 9-12 probability sample nor the probability sample of 15-year-olds. Besides the cases not in grades 9-12 tested because they belonged in our age 15 sample, there was at least one other group not in the main probability sample (grades 9-12) whom we invited to participate. This group consisted of selected schools trying one of two experimental programs in mathematics. One of the programs had been developed by Dr. Max Beberman, of the University of Illinois, and the other by Dr. E. G. Begle, who was then at Yale and is now at Stanford. The idea at the time was that the performance of students in the new program could be compared with that of appropriately matched students taking conventional mathematics courses.*

And now for the bad news!

The good news, as related above, is that 93 percent of the invited schools agreed to participate, and we were bombarded with requests to permit additional schools or groups to participate. The bad news is that the success was not uniform. We did run into trouble (or at least a profound lack of enthusiasm) in a couple of places--notably in New Jersey and California (particularly Los Angeles). The acceptance ratios among the sampled public schools in those areas were 63.6% (7 out of 11 schools) for New Jersey, 50.0% (2 out of 4) for Los Angeles, and 58.3% (14 out of 24) for the rest of California. There had been some clearly expressed opposition to Project TALENT in the state department of education in New Jersey because they viewed the project as Federal intervention. Frederick M. Raubinger, in a speech to the American Association of School Administrators early in 1959 said that Project TALENT would be "the camel's nose under the tent". Raubinger was New Jersey's Commissioner of Education at the time, and he probably was at least indirectly responsible for the comparatively poor cooperation rate in New Jersey. After all, it is not surprising if school personnel hesitate to take actions of which the state's Commissioner of Education has publicly indicated he disapproves. But the "bad news" wasn't really so bad when one considers that the opposition to Project TALENT substantially affected cooperation in only two states, and that even in those two states well over half the invited schools participated in the Project.

---

* No, this has not been done.

Identifying and contacting the corresponding junior high schools (feeder schools)

For those schools in our sample that regularly received as tenth-graders students from local junior high schools, it was necessary to identify the appropriate junior high schools, contact them, and secure their agreement to have their ninth-grade students participate in Project TALENT. This was carried out almost entirely by the Regional Coordinators, operating on the basis of instructions from Bill Gorham. (Largely because of the time pressure, most of these instructions were transmitted by telephone.) All of the invited junior high schools, as far as our records show, agreed to participate. (The fact that the senior high schools to which their students would go had already agreed to participate, and that this meant, in most cases, that the school superintendent was backing the enterprise, must have helped considerably in achieving this 100% record. Nevertheless it is a record to which we can point with pride.)

However it must be admitted that if our grade 9-12 probability sample has one weak spot it is in the ninth-grade sample, and its source is the identification of the feeder junior high schools. In the first place the instructions for defining feeder schools, having been transmitted orally, almost certainly were not interpreted uniformly by all Regional Coordinators. In the second place the dilemma of how to cope with the partial overlap between junior and senior high schools had not been faced up to in formulating the instructions to Regional Coordinators. (Some senior high schools in the sample received their tenth-grade students from several junior high schools, and some junior high schools feeding students to the selected senior high schools also sent students to senior high schools not in the sample.) Furthermore, as far as I know, a technically immaculate solution to this sampling problem had not been developed at the time, so that even if the dilemma had been faced up to in developing instructions for the Regional Coordinators it might not have been resolved in a completely satisfactory manner. A technically sound solution to the problem now exists. The author of this "History" developed it some 15 years later in connection with work on an entirely different problem, not related to Project TALENT in any way. She dubbed the solution "The 'Closed School-Cluster' Procedure for Selecting a Probability Sample of Schools". It is described in Appendix D2. (This is an instance of what is referred to in the Preface as the "HIBK" phenomenon.)

It should be clearly understood, however, that despite the fact that the messy (and in 1959 unsolved) problem created by the frequent lack of one-to-one correspondence between senior high schools and feeder junior high schools resulted in a less-than-perfect ninth-grade sample, it is nevertheless a good sample--in fact a very good sample.

But if a second Project TALENT--what we sometimes refer to facetiously as "Son-of-TALENT"--or anything resembling it is ever undertaken, it is hoped sincerely (at least by the author of this History) that the "closed school-cluster" procedure will be used, so that the sample for grade 9 will be every bit as good as those for grades 10, 11, and 12--and nobody has ever suggested that the samples for those three grades are anything less than excellent.

As has already been implied, at the time that the Regional Coordinators determined which junior high schools were feeder schools, and should therefore be invited to participate, no data were formally collected in written form concerning how many tenth-graders each senior high school in the sample received from each junior high school and how many students each of the potential "feeder schools" thus identified sent to senior high schools not in the sample. These data were needed by Marion Shaycoft, who had completed her responsibility by then of getting the test battery developed, the test booklets organized, and the five answer sheets planned, and therefore was free to turn her attention to the problem of school weights. Thus at that point data were urgently required so that there would be some reasonable basis for determining as well as could be done under the admittedly not perfect circumstances the weights to be assigned to each feeder school so that the result would be as close as could be achieved to a flawless probability sample. (Bear in mind that the closed school-cluster method didn't yet exist, so there was no way the sample or the weights could be that good, but we were unaware at that time that such a solution was waiting to be found. We recognized only that we had to do the best we could with what we had.) Under Bill Gorham's direction, therefore, an effort was made by the Regional Coordinators in areas where junior high schools existed to obtain the needed data retroactively. When these data were examined the uncomfortable suspicion was confirmed that there was one segment of the grade 9 population for whom the probability of inclusion in the sample as drawn was zero, and thus not in accordance with one of the main tenets of probability sampling, namely that every member of the population sampled must have known probability, greater than zero, of inclusion in the sample. The junior high schools with zero probability of inclusion that affected the quality of the ninth-grade sample were some of those in districts where the senior high school was selected for the sample. The junior high schools in question were the ones that did not meet whatever definition of "feeder school" had been locally operative in terms of the Regional Coordinator's operating principles.

This defect had to be corrected if at all possible. The way to do it was to identify the affected schools, select a random sample of them, and persuade them, at that late date, to participate. (By the time this particular crisis came to light it was late in the spring of 1960, and testing was almost over.) John Dailey at that point assumed personal responsibility for executing the selection of this small supplementary sample. He visited NCES, identified the junior high schools (almost 150) in the affected school districts, and then, using a table of random numbers, selected one in 20, or a total of seven. Subsequent checking against our records indicated that two of the seven were already in our sample. The remaining five constituted a technically sound probability sample of the segment of the population for which we were trying to get representation. The five junior high schools were contacted immediately, but because of the lateness of the date only two could participate. (These two schools, therefore, turned out to have extremely high weights. See Chapter 15.)

> [Author's Note
>
> As the author of this "History" I have used a fairly substantial amount of space and a fairly substantial number of words to relate the details of this particular irregularity. However, since I would not like to see its importance blown up out of all proportion it seems desirable to explain that I have gone into all this detail principally for two reasons: first, because I think future researchers should learn from our errors, and second, because to the best of my knowledge this particular irregularity has never really been documented, except in Project files.]

Figure 7-1 is the list of participating schools. Figure 7-2 is a map showing how they are distributed.

Almost all of the schools that declined to participate did so reluctantly, and only because of scheduling problems or other administrative difficulties. Therefore we decided to attempt to collect school data (i.e., response to the three school questionnaires) from them even though student data would not be secured. This effort was very successful. Almost every school that had declined to participate in the testing was entirely willing to complete the school questionnaires.

C. Administering the tests

> "There is a school in the mountains of Eastern Kentucky which can be reached only by a gravel road bounded on either side by steep hills heavily wooded with pine. Occasionally the woods open sufficiently to reveal a cabin with an outbuilding, then the trees close in again, the steep slopes silently hostile to attempts to use the land for a livelihood. At a sharp fork in the road, one turns to the right and the road dribbles out and is no more. A wide shallow brook outlines the foot of the hill directly ahead. Across the brook are a school and accompanying dormitories. The dormitories are necessary, for the isolated and rugged country makes daily trips to and from school impossible. This is the Red Bird Settlement School.
>
> "At the corner of West 205th Street and Golden Avenue in the Bronx, New York City, students gather from all over the city to study calculus and advanced courses in science. A high proportion of them will eventually earn doctoral degrees and assume key positions in science. This is the Bronx High School of Science.
>
> "Off a residential street in Washington, D.C., a driveway sweeps into a broad arc up a hill in front of the main entrance of a modern parochial secondary school. The structure is of contemporary brick-and-glass design. The school appears to convey an impression of a quiet and efficient atmosphere. Students move from room to room with a bearing which seems to reflect their pride in the physical plant itself. This is the Archbishop Carroll High School."
>
> --Design for a Study of American
> Youth, pp. 43-44

Figure 7-1. LIST OF SCHOOLS IN SAMPLE AS OF 22 JUNE 1960

ALABAMA

ALPINE—Winterboro HS
UBURN—Auburn HS
  Drake HS
IRMINGHAM—Woodlawn HS
CHILDERSBURG—Childersburg HS
FIVE POINTS—Five Points High Schools
GADSDEN—Emma Sansom HS
GREENVILLE—Greenville Tng. HS
HEFLIN—Heflin Tng. S
HUNTSVILLE—Oakwood College Acad.
JACK—Zion Chapel HS
JASPER—Walker Co. Tng. S
LAFAYETTE—Chambers Co. Tng. S
LANETT—Lanett HS
LEROY—Leroy HS
MOBILE—Dunbar JHS
PISGAH—Pisgah HS
PLANTERSVILLE—Dallas Co. HS
RED LEVEL—Red Level HS
ROCKFORD—Coosa Co. HS
SATSUMA—Satsuma HS
TUSKEGEE—Tuskegee Inst. HS
VINCENT—Vincent HS
WOODLAND—Woodland HS

ALASKA

SITKA—Sitka HS

ARIZONA

DOUGLAS—Douglas HS
  Douglas JHS
ST. MICHAEL—St. Michael's HS
THATCHER—Thatcher HS

ARKANSAS

BATESVILLE—Batesville HS
DUMAS—Reed HS
HECTOR—Hector HS
HELENA—Lake View HS
LUXORA—Luxora HS
MAGNOLIA—Magnolia HS
MARMADUKE—Marmaduke HS
MELBOURNE—Melbourne HS
MONETTE—Childress HS
MT. HOLLY—New Hope HS
MT. VERNON—Mt. Vernon HS
MOUNTAIN VIEW—Mountain View HS
PINE BLUFF—Coleman HS
  Woodrow Wilson JHS
POTTSVILLE—Pottsville HS
SMACKOVER—Smackover HS
TEXARKANA—Washington HS
WILSON—Wilson Trade HS

CALIFORNIA

ANAHEIM—Marywood HS
ARLINGTON—Divine Word Seminary
BURBANK—Burbank SHS
  John Muir JHS
CARLSBAD—Army and Navy Academy
CARMEL—Carmel HS
CENTRAL VALLEY—Central Valley HS
CHULA VISTA—Castle Brook JHS
  Chula Vista HS
  Chula Vista JHS
CROCKETT—John Swett Unified HS
DURHAM—Durham HS
FAIRFIELD—Armijo Joint-Union HS
GARDENA—Junipero Serra HS
LE GRAND—Le Grand Union HS
LOS ANGELES—Our Lady of Loretto HS
  Phineas Banning HS
  Wilmington JHS
  Woodrow Wilson HS
McARTHUR—Fall River HS
MERCED—Our Lady of Mercy S
MODESTO—Modesto Union Academy
MORGAN HILL—Live Oak Union HS
MOUNTAIN VIEW—St. Joseph's College HS
NEEDLES—Needles Union HS
PORTOLA—Portola J-SHS
EDLANDS—Cope JHS
  Redlands HS
  Redlands JHS
SACRAMENTO—St. Francis HS
SAN FRANCISCO—Mercy HS
SAN MATEO—Junipero Serra HS
SANTA CRUZ—Holy Cross HS
TORRENCE—White JHS
YUCAIPA—Yucaipa J-SHS

COLORADO

BETHUNE—Bethune HS
CROOK—Logan Co. Branch HS
DENVER—Cathedral HS
  Horace Mann JHS
  Lake JHS
  North SHS
  Skinner JHS
DOLORES—Dolores HS
JOES—Liberty HS
LAFAYETTE—Lafayette HS
MORRISON—Bear Creek HS
NUCLA—Nucla HS
STEAMBOAT SPRINGS—Steamboat Spr. HS

CONNECTICUT

COLLINSVILLE—Canton HS
HARTFORD—Northeast JHS
  Weaver HS
KENT—Kent School
LITCHFIELD—Litchfield HS
MERIDEN—Orville H. Platt HS
NEW LONDON—St. Bernard's HS
POMFRET—Pomfret School
PUTNAM—Putnam Catholic Academy
SOUTHINGTON—Southington HS

DELAWARE

BRIDGEVILLE—Bridgeville HS
DELAWARE CITY—Delaware City HS

DISTRICT OF COLUMBIA

WASHINGTON—Archbishop Carroll HS
  St. Anthony HS

FLORIDA

CENTURY—Century HS
FELLSMERE—Fellsmere HS
FT. LAUDERDALE—Central Catholic HS
GAINESVILLE—Buchholz JHS
  Gainesville HS
  Westwood JHS
GREENVILLE—Greenville HS
JACKSONVILLE—Isaiah Blocker JHS
  James W. Johnson JHS
  New Stanton HS
LAKE BUTLER—Union Co. HS
MIAMI—B. T. Washington HS
MULBERRY—Mulberry J-SHS
OCALA—Belleview-Santos HS
PALMETTO—Palmetto HS
PENSACOLA—A. V. Clubbs JHS
  Brentwood JHS
  Brownsville JHS
  Escambia HS
  Ferry Pass JHS
  Pensacola HS
  W. A. Blount JHS
SANFORD—Seminole HS
WARRINGTON—Warrington JHS
YULEE—Yulee HS

GEORGIA

ATLANTA—Sylvan Hills HS
AUGUSTA—Aquinas HS
BAINBRIDGE—Bainbridge HS
CAIRO—Washington Consolidated HS
CLAXTON—Evans Co. HS
DUBLIN—Dublin HS
FOREST PARK—Wm. A. Fountain HS
FORSYTH—Hubbard HS
FORT VALLEY—H. A. Hunt HS
GAINESVILLE—South Hall HS
HAWKINSVILLE—J. L. Bozeman Tng. Ins.
METTER—Candler Co. Tng. School
NEWNAN—Newnan-Coweta Central HS
OCILLA—Irwin Co. HS
PERRY—Houston Co. Tng. School
RICHMOND HILL—Geo. W. Carver HS
ROSSVILLE—Rossville HS
  Rossville JHS
SMYRNA—Campbell HS
SPRINGFIELD—Effingham Co. HS
VANNA—Vanna JHS
ZEBULON—Pike Co. HS

HAWAII

HONOLULU—University HS
WAHIAWA, Oahu—Leilehua High & Int. Sch.
WAILUKU, Maui—Baldwin HS

IDAHO

BOISE—Boise Junior Academy
GOODING—Gooding HS
KENDRICK—Kendrick SHS
WALLACE—Our Lady of Lourdes Academy

ILLINOIS

ABINGDON—Abingdon SHS
ALTON—Western Military Academy
ARCOLA—Arcola HS
AURORA—Marmion Military Academy
BELVIDERE—Belvidere HS
CARRIER MILLS—Carrier Mills HS
CHAMPAIGN—Champaign SHS
  Edison JHS
  Franklin JHS
CHICAGO—Amundsen HS
(See Footnote)
  Austin HS
  Carver HS
  Crane HS
  DuSable HS
  Englewood HS
  Farragut HS
  Fenger HS
  Francis W. Parker School
  Harrison HS
  Hirsch HS
  Hyde Park HS
  Kelly HS
  Kelvyn Park HS
  St. Benedict HS
  St. Patrick HS
  South Shore HS
  Steinmetz HS
  Sullivan HS
  Taft HS
  Tuley HS
  Wells HS
  Wendell Phillips HS
ELLSWORTH—Ellsworth HS
FISHER—Fisher HS
FRANKLIN PARK—East Leyden HS
FREEPORT—Aquin Central Catholic HS
GORHAM—Gorham HS
JOHNSTON CITY—Johnston City HS
LEMONT—St. Vincent DePaul Seminary
LEROY—LeRoy HS
MACOMB—Edison JHS
  Macomb SHS
MARISSA—Marissa Twp. HS
McHENRY—McHenry Com. HS
MELROSE PARK—Walther Lutheran HS
METROPOLIS—Metroplis Com. HS
MONMOUTH—Monmouth HS
MT. OLIVE—Mount Olive HS
MT. CARMEL—Mt. Carmel HS
MT. MORRIS—Mt. Morris HS
NORTHLAKE—West Leyden HS
OREGON—Oregon Com. HS
PAYSON—Seymour Com. Unit S. Dist. No. 1
RED BUD—Precious Blood Inst.
RIVERTON—Riverton HS
SESSER—Goode-Barren Twp. HS
SHANNON—Shannon Com. Unit HS
THOMPSONVILLE—Thompsonville Com. HS
VIOLA—Winola HS
  Winola JHS
WALNUT—Walnut HS
WITT—Witt HS
WOODHULL—Alwood HS
WOOD RIVER—East Alton-Wood River S
ZIEGLER—Ziegler Com. HS

INDIANA

ANDREWS—Dallas Twp. HS
BAINBRIDGE—Bainbridge HS
BATTLE GROUND—Battle Ground HS
DUNKIRK—Dunkirk HS
EAST GARY—Thos. A. Edison SHS
  Thos. A. Edison JHS
EVANSVILLE—Central HS
  Reitz Mem. HS for Girls
GOSHEN—Goshen HS
GREENSBURG—Sandcreek HS
HANNA—Hanna HS
HARTFORD CITY—Hartford City HS
INDIANAPOLIS—Arsenal Tech HS
  Avon HS
LAWRENCEBURG—Lawrenceburg Cons. HS
LOGANSPORT—Logansport HS
MADISON—Central HS
  Madison HS
MARENGO—Marengo HS
MARION—Bennett HS
MARKLE—Union Center HS
MAXWELL—Hancock Central HS
MEXICO—Mexico HS
MITCHELL—Mitchell HS
NEW AUGUSTA—Pike Twp. HS
NEW CASTLE—New Castle HS
NEW GOSHEN—Fayette Twp. HS
PLAINVILLE—Epsom HS
SELMA—Center HS

IOWA

ADEL—Adel HS
ANTHON—Anthon-Oto Ind. HS
CEDAR RAPIDS—Regis HS
CHARLES CITY—Charles City Com. HS
CHARLOTTE—Charlotte Ind. HS
CLARENCE—Clarence Com. HS
DANBURY—Danbury Catholic HS
DENISON—Denison Com. HS
DOLLIVER—Dolliver Cons. HS
DYERSVILLE—Xavier HS
ELDON—Cardinal Com. HS
GOLDFIELD—Goldfield Com. HS
GRISWOLD—Griswold Cons. HS
GUTTENBERG—St. Mary's HS
LAKOTA—Lakota Cons. HS
LAMONT—Lamont Cons. HS
LARCHWOOD—St. Mary's HS
LE CLAIRE—Le Claire Ind. School
LE MARS—Gehlen HS
LIBERTY CENTER—Southeast Warren Com. HS
MONMOUTH—Monmouth Cons. HS
MOULTON—Moulton Udell Com. S
MT. AYR—Mt. Ayr Com. HS
PAULLINA—Paullina HS
PROTIVIN—Rudolphinum HS
RENWICH—Boone Valley Com. HS
RICEVILLE—Riceville HS
TIPTON—Tipton Com. HS

KANSAS

ARKANSAS CITY—Arkansas City HS
BELOIT—St. John HS
BREWSTER—Brewster Cons. HS
COLLYER—Collyer HS
CLAY CENTER—Clay Co. Com. HS
ELK CITY—Elk City HS
EUREKA—Eureka SHS
GREAT BEND—Great Bend HS
  Harrison JHS
  Roosevelt JHS
GREENLEAF—Greenleaf HS
HOXIE—Sheridan Com. HS
IOLA—Iola HS
LAWRENCE—Central JHS
OVERLAND PARK—Kansas City Coll and Bible S
PRETTY PRAIRIE—Pretty Prairie HS
RANDOLPH—Blue Valley HS
ROSALIA—Rosalia HS
SHAWNEE—St. Joseph's HS
SPEARVILLE—Windthorst HS

KENTUCKY

ALMO—Almo HS
AUBURN—Auburn HS
AUXIER—Auxier HS
BEVERLY—Red Bird Settlement HS
BLACKEY—Letcher HS

CHICAGO—By special arrangement with the Chicago Board of Education, 1 out of every 10 students in 20 of the 38 secondary schools in the city, was randomly selected and tested. These 20 schools were selected at random.

Figure 7-1 (cont.)

CLARKSON—Clarkson HS
CORBIN—Corbin SHS
St. Camillus Academy
DANVILLE—Kentucky School for the Deaf
ELKTON—Todd Co. HS
HARDINSBURG—Breckenridge Co. HS
DERSON—Douglas HS
LEY—Maytown HS
NGTON—Sayre School
LOUISVILLE—Holy Rosary Academy
MAYSVILLE—Maysville HS
MIDDLETOWN—Eastern HS
NAZARETH—Nazareth Academy
PIKEVILLE—Pikeville HS
PLEASANT VIEW—Pleasant View HS
SOUTH PORTSMOUTH—South Portsmouth HS
TYNER—Tyner HS
WARFIELD—Warfield HS
WHITESBURG—Whitesburg HS

LOUISIANA

BATON ROUGE—Catholic HS
COLFAX—Mary E. Graham HS
COLUMBIA—Columbia HS
CONVERSE—Oak Grove HS
DONALDSONVILLE—Lowery HS
EUNICE—Charles Drew HS
FRANKLIN—Franklin HS
GILBERT—Gilbert HS
GRAND CANE—Grand Cane HS
GRANT—Fairview HS
GRAYSON—Grayson HS
HAUGHTON—Haughton HS
HOUMA—Southdown HS
LAFAYETTE—Lafayette SHS
Paul Breaux HS
LAKE CHARLES—Landry Memorial HS
NEW ORLEANS—Mid-City Baptist School
St. Joseph's HS
OAK GROVE—Oak Grove HS
SPRINGHILL—Brown HS
TALLULAH—Reuben McCall HS
THIBODAUX—Thibodaux HS
Thibodaux JHS
WATERPROOF—Waterproof HS
ZACHARY—Zachary HS

MAINE

BANGOR—Bangor HS
John Baptist HS
CHERRYFIELD—Cherryfield Academy
GORHAM—Gorham HS
MADAWASKA—Madawaska HS
SACO—Thornton Academy
SO. BRISTOL—So. Bristol HS
TENANTS HARBOR—St. George HS
YORK—York HS

MARYLAND

BALTIMORE—Booker T. Washington JHS
Chas. Hamilton Houston JHS
Fred Douglas HS
Seton HS
CATONSVILLE—St. Charles College HS
EASTON—SS. Peter and Paul HS
ELKRIDGE—Waterloo JHS
ELLICOTT CITY—Ellicott City JHS
Howard Co. SHS
FREDERICK—St. John's Literary Inst.
GREAT MILLS—Great Mills SHS
LAUREL—Pallotti HS
NEWARK—Worcester J-SHS
OLDTOWN—Oldtown J-SHS
PARKVILLE—Parkville SHS
Parkville JHS
PRESTON—Preston HS
UNION BRIDGE—Francis Scott Key J-SHS

MASSACHUSETTS

BARNSTABLE—Barnstable Trade School
BOSTON—Hyde Park HS
James P. Timilty JHS
Wm. Barton Rogers JHS
INGTON—Burlington J-SHS
CHESTNUT HILL—Beaver Country Day S
The Rivers Country Day S
DEDHAM—Ursuline Academy
DORCHESTER—Monsignor Ryan Mem. HS
HUDSON—Hudson Catholic HS
IPSWICH—Sacred Heart Juniorate
MILTON—Fontbonne Academy
NORTH ANDOVER—Brooks School
QUINCY—Quincy Trade School
ROXBURY—Mission Church HS—Girls
SAUGUS—Saugus HS
SOMERVILLE—Northeastern JHS
Somerville HS
Southern JHS
Western JHS
SWAMPSCOTT—Shaw JHS
Swampscott HS
TYNGSBORO—Academy of Notre Dame
UXBRIDGE—Uxbridge HS
W. SPRINGFIELD—W. Springfield HS
WESTON—Weston HS
WESTWOOD—Westwood HS
Westwood JHS
WEYMOUTH—Bicknell JHS
Central JHS
South JHS
Weymouth HS
Weymouth Voc. HS
WINCHENDON—Murdock HS
WORCESTER—St. John's Prep. S

MICHIGAN

ADRIAN—St. Joseph Academy
ALMA—Alma HS
BARRYTON—Barryton Rural HS
BENTON HARBOR—Benton Harbor HS
Benton Harbor JHS
BIRCH RUN—Birch Run Area HS
CARO—Caro HS
CLIMAX—Climax-Scotts Com. HS
DELTON—Delton Kellogg HS
DETROIT—Benedictine HS
Holy Redeemer HS
Mumford HS
Sacred Heart Seminary
St. Agatha HS
St. Anne's HS
DUNDEE—Dundee HS
EAST DETROIT—East Detroit HS
Grant JHS
Oakwood JHS
ELKTON—Elkton HS
ESSEXVILLE—St. John HS
GRAND RAPIDS—Godwin-Christian HS
Plymouth HS
IDA—Ida HS
IMLAY—Imlay City HS
IONIA—SS. Peter and Paul Academy
LUDINGTON—St. Simon HS
LYONS—Lyons-Muir Com. School
MARCELLUS—Marcellus HS
MASON—Mason HS
MIDDLEVILE—Thornapple Kellogg HS
OAK PARK—Oak Park SHS
PERRY—Perry HS
PETERSBURG—Summerfield HS
PEWAMO—Pewamo HS
ROSEVILLE—Sacred Heart HS
SOUTHFIELD—E. J. Lederle JHS
Southfield HS
WATERVLIET—Watervliet HS
WYANDOTTE—Our Lady of Mt. Carmel HS

MINNESOTA

AUSTIN—Pacelli HS
Queen of Angels JHS
BRAHAM—Braham HS
BROWERVILLE—Browerville HS
CANTON—Canton HS
CHISHOLM—Chisholm SHS
CIRCLE PINES—Centennial HS
COMFREY—Comfrey Public S
DULUTH—Hermantown J-SHS
EAST GRAND FORKS—Central HS
FERGUS FALLS—Fergus Falls SHS
FOREST LAKE—Forest Lake HS
FROST—Frost HS
GAYLORD—Gaylord HS
GLEN LAKE—Hennepin Co. Home S for Boys
GOODHUE—Goodhue HS
GOOD THUNDER—Good Thunder HS
HAWLEY—Hawley HS
IONA—St. Columba HS
LE CENTER—Le Center HS
MINNEAPOLIS—Minnehaha Academy
ONAMIA—Onamia HS
OWATONNA—Owatonna SHS
PAYNESVILLE—Paynesville HS
REMER—Remer HS
ROLLINGSTONE—Holy Trinity HS

MISSISSIPPI

ABERDEEN—Aberdeen HS
BROOKLYN—Forest Co. Ag. HS
CANTON—Holy Child Jesus HS
DECATUR—C. H. Boler HS
HAZLEHURST—Hazlehurst HS
LAUREL—Geo. S. Gardiner HS.
LOUISVILLE—Nanih Waiya HS
LULA—Lula-Rich Cons. HS
MIZE—Mize HS
PERKINSTON—Jim Broom Mem. HS
SANDERSVILLE—Sandersville HS
SENATOBIA—Senatobia HS
SHANNON—Shannon HS
SUMRALL—Sumrall HS
UNIVERSITY—University HS
VICKSBURG—St. Aloysius HS

MISSOURI

ALMA—Alma HS
AUGUSTA—Augusta HS
BERNIE—Bernie HS
BEVIER—Bevier HS
BOWLING GREEN—Bowling Green HS
CONCORDIA—Concordia HS
GALLATIN—Gallatin HS
GLENCOE—La Salle Institute
GOWER—Gower HS
GRANDBY—Grandby HS
GREENFIELD—Greenfield HS
INDEPENDENCE—St. Mary's HS
KANSAS CITY—East HS
Northeast HS
Northeast JHS
KINLOCH—Kinloch HS
LEE'S SUMMIT—Lee's Summit HS
LEOPOLD—Leopold R-3 HS
LIBERTY—Liberty HS
LILBOURN—Lilbourn HS
MAPLEWOOD—Maplewood-Richmond Hts. HS
OAK RIDGE—Oak Ridge HS (R-6)
ST. ANN—Pattonsville HS
Pattonsville JHS
STE. GENEVIEVE—Ste. Genevieve Public HS
ST. JOSEPH—Christian Brothers HS
ST. LOUIS—Academy of the Sacred Heart
Brittany JHS
Hanley JHS
St. Louis Prep. Seminary
Univ. City SHS
SALISBURY—Salisbury HS (R-10)
TUSCUMBIA—Tuscumbia HS
UNIONVILLE—Unionville HS
WASHBURN—Southwest HS

MONTANA

ANACONDA—Anaconda HS
BILLINGS—Billings HS
Billings West HS
Garfield JHS
Lewis and Clark JHS
Lincoln JHS
CUT BANK—Cut Bank HS
GARDINER—Gardiner HS
LAVINA—Lavina HS
LIBBY—Libby HS
THOMPSON FALLS—Thompson Falls HS

NEBRASKA

BRADY—Brady HS
CAMPBELL—Campbell HS
CRAWFORD—Crawford HS
CRESTON—Creston HS
CROFTON—Crofton HS
ELSIE—Elsie HS
GRAND ISLAND—Central Catholic HS
HEBRON—Hebron HS
LINCOLN—Lincoln HS
Whittier JHS
LOUP CITY—Loup City HS
NORTH PLATTE—St. Patrick's HS
OMAHA—Omaha Benson HS
Omaha Tech. HS
OSHKOSH—Garden Co. HS
PENDER—Pender HS
PILGER—Pilger HS
PRIMROSE—Primrose HS
WALLACE—Wallace HS

NEVADA

LAS VEGAS—Bishop Gorman HS
John C. Fremont JHS
Las Vegas HS

NEW HAMPSHIRE

ALTON—Alton HS
GREENFIELD—Crotched Mt. School
HANOVER—Hanover HS
MANCHESTER—Bishop Bradley HS
ROCHESTER—Spaulding HS

NEW JERSEY

BLACKWOOD—Mother of the Savior Sem.
BURLINGTON—St. Mary s Hall
DUNELLEN—Dunellen HS
EAST ORANGE—Ashland JHS
Clifford J. Scott HS
Columbian JHS
Franklin JHS
Garfield HS
IRVINGTON—Archbishop Walsh HS
JERSEY CITY—St. Mary HS
St. Peter's Prep. S
MONTCLAIR—Montclair Academy
NEW BRUNSWICK—MiddlesexCo. Voc. and Tech. HS
NEWARK—St. Michael HS
TENAFLY—Tenafly HS
TRENTON—Hamilton HS West
Villa Victoria Academy
WOODSTOWN—Woodstown HS

NEW MEXICO

ALBUQUERQUE—Academy for Boys
Pius X Catholic HS
BELEN—Belen SHS
JAL—Jal HS
LOGAN—Logan Municipal HS
MELROSE—Melrose HS
OJO CALIENTE—Ojo Caliente HS

NEW YORK

AMENIA—Barlow School
AMITYVILLE—Queen of the Rosary Academy
AUBURN—West HS
AU SABLE—Au Sable Forks Central Rural S
BALDWIN—Baldwin HS
Baldwin JHS
BRASHER FALLS—St. Lawrence Central S.
BUFFALO—Mt. Saint Mary Academy
CATO—Cato-Meridian Central S.
CHAMPLAIN—St. Mary's Academy
CHATEAUGAY—Chateaugay Central S.
CLYDE—Clyde Central S.
DELEVAN—Delevan-Machias Central S.
DUNKIRK—Dunkirk HS
ELBRIDGE—Elbridge Central S.
FOREST HILLS—Kew Forest S.
FORT PLAIN—Fort Plain HS
GREENE—Greene HS
GREENVILLE—Greenville Central S.
GUILDERLAND CENTER—Guilderland Central HS
HAMBURG—Hamburg Central HS
HAMILTON—Hamilton Central S.
JAMAICA—Jamaica Voc. HS
Mary Louis Academy
Thom. A. Edison Voc. HS
Woodrow Wilson Voc. HS
JOHNSON CITY—C. F. Johnson JHS
Johnson City HS
LONG ISLAND—Aviation HS
Queens Voc. HS
MANHASSET—St. Mary's HS
MORAVIA—Moravia Central HS

Figure 7-1 (cont.)

New York City

(See Footnote)

BRONX—Alexander Burger JHS, Benjamin Franklin JHS, Bronx HS of Science, Bronx [illegible] HS, Castle Hill JHS, Christopher Co[illegible]s HS, Creston JHS, David G. Farra[illegible] JHS, Dewitt Clinton HS, Elijah D. Clark JHS, Elizabeth Browning JHS, Evander Childs HS, Frank D. Whalen JHS, Grace H. Dodge Voc. HS, Henry J. Bruckner HS, Henry Hudson JHS, Herman Ridder JHS, James M. Kieran JHS,, James Monroe HS, Jane Addams Voc. HS, John Dwyer JHS, John Philip Sousa JHS, John Peter Tetard JHS, Jordan L. Mott, JHS, Joseph H. Wade JHS, Macombs JHS, Morris HS, Mosholu Parkway JHS, Olinville JHS, Paul Laurence Dunbar JHS, Paul Hoffman JHS, Prospect JHS, Riverdale JHS, St. Helena's HS, Sam. Gompers Voc. and Tech HS, Theodore Roosevelt HS, Thomas Knowlton JHS, Walton HS, Wm. H. Taft HS, Wm. W. Niles JHS

BROOKLYN—Abraham Lincoln HS, Abraham Lincoln JHS, Alexander Hamilton Voc. HS, Andries Hudde JHS, Arthur S. Somers JHS, Bay Ridge HS, Bensonhurst JHS, Berriman JHS, Bishop Loughlin Mem. HS, Boys HS, Brooklyn HS of Automotive Trades, Brooklyn Tech. S., Bushwick HS, Charles O. Dewey JHS, Clara Barton Voc. HS, Col. David Marcus JHS, David A. Boody JHS, Decatur JHS, East N. Y. Voc. HS, East N. Y. JHS, Eastern District HS, Edward B. Shallow JHS, Eli Whitney Voc. HS, Erasmus Hall HS, F. K. Lane HS, Francis Scott Key JHS, Ft. Hamilton HS, Georgu Gershwin JHS, Geo. W. Wingate HS, Geo. Westington Voc. and Tech. HS, Girls HS, Halsey JHS, James Madison HS, John D. Wells JHS, John Ericsson JHS, John Jay HS, John J. Pershing JHS, John M. Harrigan JHS, John Marshall JHS, John Wilson JHS, Lafayette HS, Lefferts JHS, Mark Twain JHS, Marine Park JHS, Mark Hopkins JHS, Meyer Levin JHS, Michael Friedsam JHS, Midwood HS, Montauk JHS, Nathan Hale JHS, Nathaniel Macon JHS, New Utrech HS, Prospect Heights HS, Sam J. Tilden HS, Sands JHS, Sarah J. Hale Voc. HS, Seth Low JHS, Sheepshead Bay HS, Shell Bank JHS, St. Clair McKelway JHS, Stranahan JHS, The Ditmas JHS, The Peter Rouget JHS, Thos. Jefferson HS, W. Arthur Cunningham JHS, Walt Whitman JHS, Whitelaw Reid JHS, Wm. Alexander JHS, Wm. E. Grady Voc. HS, Wm. J. Gaynor JHS, Wm. J. Morrison JHS, Wm. H. Maxwell Voc. HS, Wm. McKinley JHS, Willoughby JHS, Winthrop JHS

MANHATTAN—B.D.L. Southerland JHS, Benj. Franklin HS, Booker T. Washington JHS, C. Sumner JHS, Central Coml. HS, Chas. E. Hughes HS, Chelsea Voc. HS, Corlears JHS, Edward W. Stitt JHS, Food Trades Voc. HS, Frederick Douglas JHS, Geo. Washington HS, Gustave Straubenmuller JHS, Haaren HS, Harriet B. Stowe JHS, HS of Commerce, HS of Music and Arts, Humboldt JHS, Inwood JHS, James Otis JHS, Jefferson Park JHS, J. Fenimore Cooper JHS, Julia Richman HS, Joan of Arc JHS, John S. Roberts JHS, La Salle JHS, Loyola School, Mabel Dean Bacon Voc. HS, Machine & Metal Tr. Tech & Voc. HS, Manhattanville JHS, Margaret Knox JHS, Metropolitan Voc. HS, N. Y. School of Printing, Ottilie M. Beha JHS, Rhodes Prep. School, Rice HS, Robert E. Simon JHS, Robert F. Wagner JHS, School for Deaf JHS, Seward Park HS, Simon Baruch JHS, Stuyvesant HS, The HS of Art and Design, The HS of Fashion Industries, Wadleigh JHS, Wash. Irving HS, Wm. L. [illegible]ger JHS, Wm. J. O'Shea JHS, York[illegible] Voc HS

[illegible]EENS—Andrew Jackson HS, Astoria JHS, Bayside HS, Ben. N. Cardozo JHS, Corona JHS, Daniel Carter Beard JHS, East Elmhurst JHS, Edgar D. Shimer JHS, Edward Bleeker JHS, Far Rockaway HS, Flushing HS, Forest Hills HS, George J. Ryan JHS, Grover Cleveland HS, Harold G. Campbell JHS, Hillside JHS, Horace Greeley JHS, Jamaica HS, John Adams HS, Joseph Pulitzer JHS, Long Island City HS, Louis Pasteur JHS, Marie Curie JHS, Martin Van Buren HS, Nathaniel Hawthorne JHS, Newtown HS, Oliver W. Holmes JHS, Parsons JHS, Queens Village JHS, Richmond Hill HS, Robert A. Van Wyck JHS, Rockaway Beach JHS, Russell Sage JHS, Sam. Huntington JHS, Springfield Gardens JHS, Steinway JHS, Steven A. Halsey JHS, Wm. C. Bryant HS, Wm. H. Carr JHS, Wm. Cowper JHS, Woodside JHS

RICHMOND—Curtis HS, Edwin Markham JHS, New Dorp HS, Port Richmond HS, Ralph R. McKee Voc. Tech HS, Tottenville HS

NIAGARA FALLS—Madonna HS
Trott Voc. HS
OSSINING—Mary Immaculate S.
OTEGO—Otego Central HS
PATCHOGUE—Seton Hall HS
PHILADELPHIA—Indian River Central HS
PT. WASHINGTON—Carrie B. Weber JHS
John Philip Sousa JHS
Paul D. Schrieber HS
ROCHESTER—St. Andrews Seminary
SAVANNAH—Savannah HS
SCHOHARIE—Schoharie Central School
STAATSBURG-ON-HUDSON—The Anderson School
SYRACUSE—Cathedral Academy
TARRYTOWN—Our Lady of Victory Academy
WALLKILL—Wallkill Central HS
WARSAW—Warsaw HS
WESTPORT—Westport Central HS
YONKERS—Elizabeth Seton School

NORTH CAROLINA

ALBEMARLE—Endy HS
ASHEVILLE—Acad. of St. Genevieve-of-the-Pines
Biltmore HS
Plonk S. of Creative Arts
BELMONT—Belmont HS
CAMERON—Cameron HS
CHADBOURN—Chadbourn HS
CLAYTON—Cleveland HS
DENTON—Denton HS
GASTONIA—Highland HS
GREENSBORO—Bessemer HS
Page HS
HILLSBORO—Hillsboro HS
LENOIR—Freedman HS
Kings Creek HS
LOUISBURG—Riverside HS
LUMBERTON—Lumberton HS
MARSHVILLE—Marshville HS
MONROE—Union HS
Walter Bickett HS
NEWTON GROVE—Hobbton HS
NORLINA—Norlina HS
PINE HALL—Pine Hall HS
PRINCETON—Princeton HS
RALEIGH—J. W. Ligon HS
ROPER—Washington Co. Un.HS
SHALLOTE—Union HS
SKYLAND—Valley Springs HS
SMITHFIELD—Smithfield HS
STANFIELD—Ridgecrest HS
SUMMERFIELD—Laughlin HS
TROY—Peabody HS
VALE—Union HS
WALNUT—Walnut HS
WASHINGTON—P. S. Jones HS
WINGATE—Wingate College HS

NORTH DAKOTA

FORBES—Forbes HS
GLENFIELD—Glenfield Public HS
HANKINSON—Hankinson HS
HAYNES—Haynes HS
KLOTEN—Kloten HS
PLAZA—Plaza HS
WAHPETON—Wahpeton Central HS

OHIO

AKRON—East HS
Goodyear JHS
BLOOMDALE—Elmwood HS
CANTON—J. H. Lehman HS
CARROLLTON—Carrollton HS
CENTERBURG—Centerburg Local HS
CINCINNATI—Regina HS
CLEVELAND—Cuyahoga Hts. HS
Holy Name HS
Mayfield HS
St. Peter HS
COALTON—Coalton Local HS
COLUMBUS—Coll. of St. Chas. Borromeo
Hamilton Twp. HS
CONESVILLE—Conesville HS
COSHOCTON—Coshocton HS
DOVER—St. Joseph's HS
DOYLESTOWN—Chippewa HS
EUCLID—Euclid JHS
Euclid SHS
Shore JHS
FRANKFORT—Frankfort HS
GETTYSBURG—Gettysburg HS
GREEN CAMP—Green Camp HS
GREEN SPRINGS—Green Springs HS
GREENHILLS—Greenhills HS
HAMILTON—Hamilton Catholic HS
HARTVILLE—Lake HS
HOWARD—Howard HS
HURON—Huron HS
KINGSTON—Saltcreek Twp.HS
LEAVITTSBURG—Warren Twp. Local HS
LEES CREEK—Simon Kenton HS
LODI—Cloverleaf HS
LUCASVILLE—Valley Local HS
MANSFIELD—Mansfield SHS
NEW STRAITSVILLE—New Straitsville HS
NEWARK—Newark Catholic HS
NORTH CANTON—Hoover HS
N. ROYALTON—N. Royalton HS
OSTRANDER—Scioto Valley HS
PAINESVILLE—Riverside HS
PEEBLES—Peebles Local HS
RICHWOOD—Richwood HS
SANDUSKY—St. Mary HS
SHAKER HEIGHTS—Laurel HS
STRONGSVILLE—Strongsville HS
TOLEDO—Ottawa Hills HS
TONTOGANY—Otsego HS
WADSWORTH—Wadsworth HS
W. JEFFERSON—Jefferson Local HS
W. LAFAYETTE—W. Lafayette HS
YOUNGSTOWN—Chaney HS
West JHS

OKLAHOMA

ALLEN—Allen HS
Allen JHS
ANADARKO—Riverside Indian School
CHEROKEE—Lambert HS
CHICKASHA—Friend HS
COMMERCE—Commerce HS
DALE—Dale HS
DOVER—Dover HS
ENID—Emerson JHS
Enid HS
Longfellow JHS
Waller JHS
FREDERICK—Boyd HS
GUYMON—Guymon HS
Guymon JHS
KANSAS—Kansas J-SHS
MARLAND—Marland HS
NOWATA—Nowata HS
Nowata JHS
OKLAHOMA CITY—Putnam City HS
OOLOGAH—Oologah HS
PURCELL—Purcell HS
RAVIA—Ravia HS
SEMINOLE—Varnum HS
SHAWNEE—St. Gregory's HS
THOMAS—Thomas J—SHS
TULSA—Bell JHS
Cleveland JHS
Hamilton JHS
Will Rogers HS
Wilson JHS
VANOSS—Vanoss Public School
WILBERTON—Wilberton HS

OREGON

COLTON—Colton HS
EUGENE—Willamette HS
GERVAIS—Gervais Union HS
MEDFORD—Rogue River Jr. Acad.
PENDLETON—Pendleton HS
Pendleton JHS
PORTLAND—Franklin HS
Washington HS
SALEM—Sacred Heart Academy
SISTERS—Sisters HS
VERNONIA—Vernonia HS
WALDPORT—Waldport HS

PENNSYLVANIA

ALEXANDRIA—Juniata Valley HS
BARNSBORO—Barnsboro HS
BEAVER FALLS—Beaver Falls Area HS
Beaver Falls Area JHS
BROCKWAY—Brockway Area HS
BROOKVILLE—Brookville Area Jt. HS
CARNEGIE—Carnegie HS
Harding JHS
CLARION—Clarion Jt. J-SHS
CLIFTON HTS—Clifton Heights HS
COALPORT—Beccaria-Coalport-Irvona Jt. SHS
COLUMBIA—St. Joseph's Academy
COMMODORE—Purchase Line Jt. HS
COOPERSBURG—Southern Lehigh Union HS
DALLASTOWN—Dallastown Area Jt. HS
E. SMITHFIELD—Smithfield-Ridgeby-Ulster Jt. HS
EDDYSTONE—Eddystone SHS
ERIE—Lawrence Pk. Twp. HS
EVERETT—Everett-Southern Jt. HS
FISHERTOWN—Chestnut Ridge Jt. HS
HAMBURG—Hamburg HS
HOUSTON—Chartiers Houston Jt. HS
KINGSTON—Wyoming Seminary
LANCASTER—Edward Hand JHS
J. P. McCaskey HS
J. Reynolds JHS
Lancaster Twp. JHS
LUCINDA—St. Joseph's HS
MAHANOY CITY—Mahanoy Area HS
MONACA—Monaca Boro SHS
MONROEVILLE—Gateway SHS
Monroeville JHS
NEFFSVILLE—Mannheim Twp. HS
N. BRADDOCK—Scott SHS
NORTH EAST—St. Mary's College
OAKMONT—Oakmont Boro HS
PENFIELD—Huston Twp. HS
PHILADELPHIA—The Wm. Penn HS
So. Phila. HS
PITCAIRN—Pitcairn JHS
PITTSBURGH—St. Elizabeth HS
So. Hills Catholic HS
So. Voc. HS
Westinghouse HS
RIDLEY PARK—Ridley Pk. HS
ROGERSVILLE—W. Greene HS
SCHUYLKILL HAVEN—Schuylkill Haven Area Jt. HS
SCRANTON—St. Paul HS
SHAMOKIN—Shamokin Area Jt.HS
SPANGLER—Northern Cambria Jt. HS
SWIFTWATER—Pocono Mt. Jt. HS
TOWNVILLE—Townville Jt. HS
VERONA—Verona HS
WILKES-BARRE—Benjamin Franklin JHS
Marymount HS
Hanover Twp. Mem. HS

RHODE ISLAND

PROVIDENCE—Central SHS
Gilbert Stuart JHS
WAKEFIELD—S. Kingstown J-SHS

SOUTH CAROLINA

ABBEVILLE—Wright HS
BLACKVILLE—Blackville HS
CLOVER—Clover HS
DARLINGTON—St. John's HS

NEW YORK CITY—By special arrangement with the Board of Education, 1 out of every 12 students in every secondary school in the city was selected at random and tested.

Figure 7-1 (cont.)

EASTOVER—Webber HS
GREELEYVILLE—County Tr. S
GREENWOOD—Greenwood HS
Northside JHS
Southside JHS
LANCASTER—Lancaster SHS
MADISON—Cleveland HS
MARION—Marion HS
SUMTER—Lincoln HS
Mayewood HS
TAMASSEE—Tamassee DAR HS
WELLFORD—Florence Chapel HS
WHITMIRE—Whitmire HS

SOUTH DAKOTA

BERESFORD—Beresford HS
BRANDT—Brandt HS
DOLAND—Doland HS
MARTIN—Bennett Co. HS
NEW HOLLAND—Dakota Christian HS
RAPID CITY—Cathedral HS
REDFIELD—Redfield HS
SCOTLAND—Scotland HS
SELBY—Selby HS
VIBORG—Viborg HS
WITTEN—Witten HS

TENNESSEE

BEECH BLUFF—Beech Bluff HS
BULLS GAP—Bulls Gap HS
CHATTANOOGA—The McCallie School
DANDRIDGE—Maury HS
DUNLAP—Sequatchie Co. HS
FOUNTAIN CITY—Halls HS
GOODLETTSVILLE—Goodlettsville HS
HARRIMAN—S. Harriman HS

KNOX COUNTY*—CONCORD—Blue Grass School, Concord School, Farragut HS, Hardin School; CORRYTON—Corryton School, Gibbs School, Skaggston School; KNOXVILLE—Alice Bell School, Amherst School, Anderson School, Arminda School, Asbury School, Austin HS, Bell Camp School, Bearden School, Bearden HS, Beardsley JHS, Bonnie Kate School, Cedar Grove School, Central HS, Christenberry JHS, Dante Elementary, Dora Kennedy School, East HS, Fairview School, First Lutheran School, Fountain City School, Fulton HS, Galbraith School, Gap Creek School, Heiskell Elem. School, High Bluff School, Holston HS, Holy Ghost School, Inskip School, John Sevier School, Karns Elem. School, Karns HS, Klondike School, Knoxville Catholic HS, Lyons View School, Marbledale School, Mooreland Heights School, Mount Olive School, New Hopewell School, Norwood School, Park JHS, Pleasant Ridge School, Pond Gap School; Ramsey School, Ridgedale School, Ritta School, Riverdale School, Robert Huff School, Rocky Hill School, Rule J-SHS, Sacred Heart School, Sevier Home School, Shannondale School, Simpson School, Seventh Day Adventist Schools, Smithwood School, South J-SHS, Spring Hill School, St. Mary's Elem. School, Sterchi School, Sunnyview School, Sunrise School, Tyson JHS, Union Elem. School, Vestal School, Vine JHS, Webb School of Knoxville, West Haven School, West HS, West Hills School, White School, Wood Hill School, Young HS; MASCOT—Clear Springs School, Mascot School; POWELL—Green Hill School, Powell Elementary School, Powell HS; STRAWBERRY PLAINS—Carter Elem. School, Carter HS, Midway School, Smith School.

LENOIR CITY—Lenoir City HS
LOUDON—Loudon County HS
LYNVILLE—Jones HS
McMINNVILLE—McMinnville City HS
RAMER—Ramer HS
SHARON—Sharon HS
SPEEDWELL—Powell Valley HS
TIPTONVILLE—Tiptonville HS
TRIMBLE—Trimble City HS
TYNER—Tyner HS and JHS
WOODBURY—Central HS

TEXAS

ARCHER CITY—Archer City HS
ANTON—Anton HS
ASPERMONT—Aspermont HS
BRIDGEPORT—Bridgeport HS
BEAUMONT—St. Anthony's HS
CANADIAN—Canadian HS
CHANNELVIEW—Channelview HS
CLEVELAND—Tarkington HS
DALLAS—Bryan Adams HS
E. H. Cary JHS
Thomas Jefferson HS
DETROIT—Detroit HS
EL CAMPO—EL Campo HS
EL PASO—Lydia Patterson Inst.
FLUVANNA—Fluvanna HS
FORT WORTH—C. F. Brewer HS
C. F. Brewer JHS
GAINESVILLE—Gainesville HS
GALVESTON—Ursuline Academy
GARRISON—Garrison High Schools
GRAND SALINE—Grand Saline HS
GREGGTON—Pine Tree HS
Pine Tree JHS
HAMSHIRE—Hamshire HS
HARDIN—Hardin HS
HONDO—Hondo HS
IRA—Ira HS
LA VERNIA—La Vernia HS
LEONA—Leona HS
LITTLEFIELD—Littlefield HS
LUBBOCK—Carrol Thompson JHS
Dunbar J-SHS
J. T. Hutchinson JHS
Monterey SHS
O. L. Slaton JHS
R. W. Matthews JHS
S. C. Wilson JHS
Tom S. Lubbock SHS
W. B. Atkins JHS
MAPLE—Three Way HS
MINERAL WELLS—Mineral Wells HS
NAVASOTA—Navasota HS
NEEDVILLE—Needville HS
OVERTON—Leveretts Chapel HS
PEARSALL—Pearsall HS
PORT ARTHUR—Sacred Heart HS
Thomas Edison JHS
Thomas Jefferson HS
Woodrow Wilson JHS
QUEEN CITY—Queen City HS
RAYMONDVILLE—Raymondville HS
REDWATER—Redwater HS
SAN ANTONIO—Douglas JHS
Horace Mann JHS
Longfellow JHS
Peacock Mil. Academy
Thomas Jefferson HS
SEAGOVILLE—Seagoville HS
S. HOUSTON—S. Houston HS
S. Houston JHS
STINNETT—Stinnett HS
TEMPLE—Academy HS
TRENT—Trent HS
TUSCOLA—Jim Ned HS
VAN—Van HS
WACO—Catholic HS
WAXAHACHIE—Turner HS
WEIMAR—Karl Downs HS
WEINERT—Weinert HS

UTAH

CIRCLEVILLE—Piute HS
LEWISTON—Lewiston JHS
MONROE—South Sevier HS
South Sevier JHS
MT. PLEASANT—Wasatch Academy
OREM—Lincoln JHS
Orem HS
RICHMOND—No. Cache HS
SMITHFIELD—Smithfield JHS

VERMONT

MANCHESTER—Burr and Burton Seminary
MILTON—Milton HS
MONTPELIER—St. Michael's HS
NEWBURY—Newbury HS
WATERBURY—Waterbury HS

VIRGINIA

ALEXANDRIA—Geo. Washington HS
AMELIA COURT HOUSE—Amelia HS
CHARLOTTE COURT HOUSE—Randolph Henry HS
EASTVILLE—Northampton HS
FIELDALE—Fieldale HS
GOODVIEW—Stewartsville HS
KING WILLIAM—King William HS
LYNN HAVEN—Princess Anne HS
NORA—Ervinton HS
NORFOLK—Grandby HS
PORTSMOUTH—St. Paul's Central HS
PULASKI—Pulaski HS
STANLEY—Stanley HS
STAUNTON—Staunton Military Academy
TAPPAHANNOCK—Essex Co. HS
WEST POINT—West Point HS

WASHINGTON

EAST STANWOOD—Cedar Home SDA School
KIRKLAND—Kirkland JHS
Lake Washington SHS
LAKE STEVENS—Lake Stevens HS
LANGLEY—Langley HS
MONROE—Monroe HS
OAKVILLE—Oakville HS
PUYALLUP—Puyallup HS
REDMOND—Redmond JHS
ST. JOHN—St. John HS
SEATTLE—Ballard HS
Monroe JHS
Whitman JHS
SPOKANE—Bishop C. D. White Prep. S
Gonzaga Prep. S
SPANGLE—Upper Columbia Academy
WILBUR—Wilbur HS
YAKIMA—A. C. Davis SHS
Franklin JHS
Washington JHS

WEST VIRGINIA

CHARLES TOWN—Charles Town HS
ELIZABETH—Wirt Co. HS
KINGWOOD—Kingwood HS
MONTCOAL—Marsh Fork HS
MONTGOMERY—Montgomery HS
PARKERSBURG—Parkersburg Catholic HS
PINE GROVE—Pine Grove HS
QUINWOOD—Crichton HS
SMOOT—Smoot HS
WAR—Big Creek HS
WHEELING—Warwood HS
WILLIAMSON—Williamson HS

WISCONSIN

BEAVER DAM—Beaver Dam SHS
CLEAR LAKE—Jt. HS, Dist. No. 1
DURAND—Durand HS
GREEN BAY—Our Lady of Charity HS
KENOSHA—Kemper Hall
LADYSMITH—Ladysmith HS
LODI—Lodi HS
MADISON—Madison East SHS
MARKESAN—Markesan HS
MILWAUKEE—Kosciuszko JHS
Mercy HS
No. Div. HS
St. John's Cathedral HS
So. Div. HS
Walker JHS
MINERAL POINT—Mineral Point HS
OAKFIELD—Oakfield HS
ONTARIO—Ontario HS
PRAIRIE DU SAC—Prairie Du Sac HS
PRESCOTT—Prescott HS
ROSENDALE—Union HS
SALEM—Central Union HS
SOLON SPRINGS—St. Croix HS
TREMPEALEAU—Trempealeau HS
WEST ALLIS—Horace Mann JHS
W. Milwaukee HS

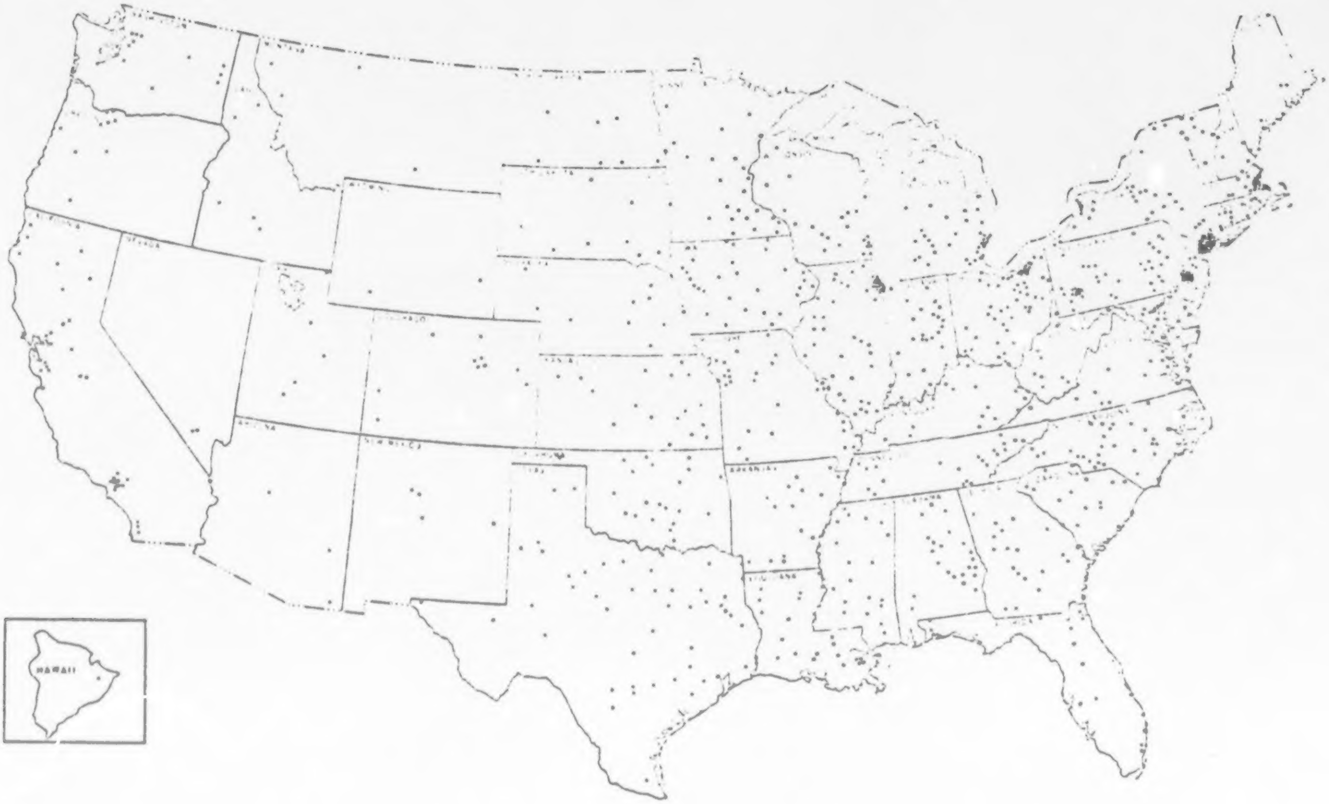
WYOMING

GREEN RIVER—Green River HS
SUNRISE—Sunrise HS

---

* KNOX COUNTY—For a special regional study, all students in grades 8-12 were tested in all schools in Knox County and Knoxville.

Figure 7-2. THE NATIONAL SAMPLE OF SCHOOLS



All three of the schools mentioned in the above quotation were in Project TALENT. They are mentioned only to give some notion of the tremendous diversity among the schools included. This diversity necessitated considerable flexibility in arrangements for test administration. In each participating school a Local Coordinator was appointed (usually by the principal in consultation with the Regional Coordinator). This Local Coordinator, working under the direction of the Regional Coordinator, trained the teachers who would actually administer the tests. He (or she) was also responsible for receiving and inventorying the test materials, distributing them to the teachers, collecting them after the testing sessions were completed, supervising the necessary checking operations, and shipping the answer sheets to Measurement Research Center (MRC), in Iowa City, where they would be processed.

The test battery required two full days, but there was considerable flexibility in how the requisite four half-days might be arranged. For instance in some schools the tests were administered during four consecutive mornings.

Each student in the sample was given--and asked to keep--an identification card (the stub of the IBM card on which his six-digit identification number was printed) identifying him as a member of the Project TALENT sample. The TALENT identification card bore, in addition to the six-digit number and the student's name (which he filled in himself) the mailing address of Project TALENT, and a request that sample members keep us informed of address changes and name changes. The students seemed to like these ID cards. (Several years later, a sample member wrote us a most appreciative letter; he had lost his wallet in a movie theater and it was returned to him because it contained something identifying him--his Project TALENT identification card!)

There had been some concern that perhaps after the first day of testing a lot of the students would decide to be "sick" the second day, and not show up. We needn't have worried. Overall attendance was a little better than average during the days of the tests, and in many schools it was noticeably better the second day than the first!

The Local Coordinators were requested to arrange make-up sessions for absentees if it was at all feasible to do so. Although in some schools this proved not to be possible, many schools did manage to have make-up sessions so that those who had been absent on one or both days could be tested.

D. E. F. Lindquist and Measurement Research Center

In the early 1950's, before Project TALENT was begun, the most advanced general-purpose scoring machine available was the IBM-805. This machine, which operated electrically (not electronically), transmitted current electrographically (i.e., across the graphite pencil marks made with a very soft lead pencil). It could score only one side of the answer sheet at a time. A second insertion was required for the second side. And both insertions were manual operations. There was no automatic feeding--even though IBM had perfected automatic feeding for IBM cards decades earlier. The machine could handle up to 150 five-choice items per answer-sheet side and up to three test scores or part scores and a total in a single insertion. Scoring was of the analogue type, not digital; the score was registered by a needle. The operator had to determine which integer the needle was closest to, and write it (with his or her right hand*) on a projecting edge of the answer sheet. An experienced operator could score a few hundred answer sheets in an hour. These scoring machines have not yet gone the way of the dinosaurs. Some still exist and are still operating--but in the early 1950's they were the only thing generally available, and the world of educational research was full of them.

Imagine, then, the excitement with which E. F. Lindquist's announcement, at the ETS Invitational Conference on Testing Problems in 1953, was greeted. He announced that he was building a scoring machine that would automatically feed and score 6,000 answer sheets an hour, would score both sides at once, and would yield as many as 14 scores, either raw or corrected for chance, on a single insertion. The 8½" x 11" answer sheet would hold roughly 1,000 five-choice items. The machine would operate via optical scanning (an electronic process) rather than via electrographic conduction. The output, too, would be vastly superior to anything then available.

When Lindquist made his announcement in 1953, he made it perfectly clear that the scoring machine was just in the planning stage, and that while the components all existed, and worked individually, they had not yet been put together and tested in combination.** He implied therefore that changes in design were not unlikely and that it would not be surprising if the scoring machine, when built, turned out to be somewhat different in operation from what was planned. It did turn out to be different, in some ways that were extremely advantageous and in other ways that proved somewhat inconvenient. The major improvement, as things turned out, was that miraculously, the direct output was neither answer sheets with the scores recorded on them nor printouts, but ready-punched IBM cards with the examinee's name, other identifying information, and test scores already on them!

---

* Individuals lucky enough to be left-handed were spared the "privilege" of operating this horse-and-buggy scoring machine.

** Lindquist made it clear that the design of this new scoring machine would be based, in large part, on developmental work done previously by Phillip J. Rulon.

Thus not only did it take only a tiny fraction of the time required by the old 805 machines, but it was far more accurate and, perhaps best of all, it eliminated the time-consuming, costly, error-producing process of key-punching.

Another improvement, of considerably less import but advantageous nonetheless, was that instead of having space for under 1,000 five-choice items, the answer sheet could hold (and the Scoring Machine could cope with) nearly 1,200.

As for the changes in the opposite (i.e., disadvantageous) direction,their impact was weaker. It turned out that only 13 scores, not 14, could be obtained on a single insertion. Somewhat more inconveniently, it turned out that for each test for which scores corrected for chance were desired, two separate scores--the number of items answered correctly and the number attempted--would have to be obtained, thus being likely to double the number of insertions required, and thus double the cost, if more than six tests were involved for which scores corrected for chance were desired. The scoring machine at that time was in a constant state of flux, continually being modified in some way by MRC's engineers, to add a new feature or improve it some other way. We were already drawing up final specifications for MRC processing of answer sheets A and C on the assumption that the earlier machine specs, according to which a set of scores corrected for chance would require no more insertions than uncorrected scores, applied, when we found out about the change.

We had wanted most of the scores "corrected for chance", alias "corrected for guessing". Note to reader: Yes, we were fully aware that "corrected for chance" and "corrected for guessing" are to some degree overstatements. Of course the correction isn't perfect. But we felt strongly that a simple correction (by the conventional formula) even though admittedly not perfect, was far better than the alternative (no correction). Unfortunately, however, the change in Scoring Machine specifications meant that the "simple correction" would not be so simple to achieve. Doubling the number of answer sheet insertions would double the costs. And there was no way that the project budget could absorb an increase of that magnitude. It was clear that we would have to settle for "number right" scores except for the four very highly speeded tests* of various kinds of computational and perceptual speed and accuracy, for which it would be possible to get counts of both the number right and the number attempted. That would permit us to get formula scores which were deliberately set to overcorrect for wrong answers--a legitimate technique (indeed, a necessary technique) for that sort of test with the sort of predictive use for which those tests were intended. Settling for having formula scores for only four tests was a compromise, but it was a compromise dictated by reality. And it wasn't a bad compromise. Number-right scores have a very high correlation with "corrected" scores. But even a very high correlation allows an occasional "outlier" case and a resultant injustice or serious misevaluation. We are aware, of course, that many otherwise admirable people think "number right" is the only way to go. However this "History" probably isn't the proper forum to debate the matter,** so we shall go on to something else, after contenting ourselves with pointing out that this compromise did not mean that we were abandoning for all time the idea of "corrected" scoring. (See Chapter 12.)

Before dropping (at least temporarily) the matter of Scoring Machine limitations we should mention one further problem--one which was to complicate matters considerably for us (and as a matter of fact still is causing complications. See Chapter 13.) That was the problem of distinguishing between legitimate zero scores (failure to answer even one item correctly) and blanks or "missing data" (i.e., failure to take the test at all). The Scoring Machine couldn't make this distinction. Missing data, like legitimate zero scores, were represented on the IBM card output by zero punches.

When the new scoring machine had been built and had been in use for a couple of years. Lindquist undertook to build another machine, even more advanced, whose card output would provide item responses, rather than test scores. This machine was called the "document reader" (as opposed to the "scoring machine"). Each of the 80 card columns of the IBM card output could register any one of 39 response codes: the 26 letters of the alphabet, the 10 digits, the two extra "overpunch" positions, or a blank. Each of the 80 39-character positions on the answer sheet could be reorganized to contain as many as six five- or six-choice items. This meant that even after allowing answer spaces for name and other identifying information it would still be possible to get well over 400 items on an answer sheet, some with as many as six choices. And in the case of items containing no more than six options each, it would be possible to double-punch the output cards so that each card column would contain responses to two items. (Items with more than six choices would also be possible, and equally easy to process, but each such item would cut down on the total number of items that would fit on the answer sheet; and since these items would preclude double-punching they would also reduce the number that would fit on an output card.)

---

* These four tests are:
  1. Arithmetic Computation
  2. Table Reading
  3. Clerical Checking
  4. Object Inspection

** The author has already expressed her ideas on the subject in "Guide to the Development, Evaluation, and Use of Criterion-Referenced Tests", (1976) pp. 81-83.

When John Flanagan was developing the outlines of the plan for Project TALENT, early in 1958, he visited Dr. Lindquist at MRC to find out about the new scoring machine--how effective it was, how feasible it would be to handle Project TALENT's scoring on it, and in general to explore the possibilities. On that visit he learned about the Document Reader, which was then just in the blueprint stage; construction had not yet begun. It was immediately apparent from the description that the Document Reader would be a key element in Project TALENT. This incipient machine would have one major advantage over the Scoring Machine; it would provide greater flexibility since items could be combined and recombined in different ways, and different scoring formulas could be experimented with, if scoring were postponed and done later by computer, rather than being handled on the Scoring Machine.

As a matter of fact some of the most crucial features of Project TALENT--specifically the nearly 400 items of background information that constituted the Student Information Blank--would have been entirely out of the question if the Document Reader had not been promised.

In 1959 Lin was able to assure John Flanagan that the Document Reader would be built and operating in time to handle those aspects of the Project TALENT data processing for which it was appropriate. The Scoring Machine, too, would be available. Because of its newness and its innovative character, problems--many problems--had been encountered in using the scoring machine. Lin told John that over a fairly long period of time there had been so much trouble with the scoring machine, and so much "down time", that it had only been usable about four hours a month. But--and this is a very important "but"--with just those four hours a month of usable time, it was possible to process the entire Iowa Every-Pupil Testing Program and keep up with all the many other testing programs using their services. That bit of substantive information, together with Lin's assurances that both machines would be available, functioning, and ready to process the projected TALENT data, provided a basis for firmer planning. Ideally, it would have been best to have all the answer sheets processed on the Document Reader. But as the plans developed and grew firmer, it became obvious that that would not be feasible. One important reason was that Lin could not promise that the Document Reader would be ready and working before the fall of 1960. But it was going to be necessary to get reports of results to participating schools as soon as possible. The only machine that would be available at a time that would make this possible was the Scoring Machine. Therefore, as a compromise, we decided to put the tests on which we would need scores right away on two answer sheets (Answer Sheets "A" and "C") that would go through the Scoring Machine. The remaining data--mostly responses to questionnaires (specifically the Student Information Blank and the Interest Inventory) but also responses to about 36 percent of the Information test items--would go on two answer sheets (B1 and B2) that would be processed on the Document Reader. A fifth document, "Record Form Z", which would also be processed on the Document Reader, was designed to contain identifying information--name, address, sex, grade, age, and date of birth--as well as the six-digit student identification number used to link all the answer sheets for the same student. (Enough identifying information would also appear on Answer Sheets A and C, however, to permit students' scores on selected tests to be reported to the schools.)

> [Author's Note (again):
>
> A good deal of space has been devoted to a discussion of problems and difficulties that arose in connection with use of MRC's Scoring Machine and Document Reader, and I do not want to leave the wrong impression. Even though by today's standards both machines were primitive, in that they lacked a direct document-to-tape capability, and required, instead, an intermediate step involving the production of IBM cards, nevertheless in 1960 these two machines were in the vanguard of advanced technology. With them, processing the TALENT data was nightmarishly complex, but without them the project would have been impossible. (It couldn't have been done just one year earlier, because the Document Reader would not have existed.)]

E. Implementation

1. Before MRC (The "TALENT Laundry")

Answer sheets were shipped from each school direct to MRC by the Local Coordinator. Before they could be passed through the Document Reader or Scoring Machine they had to be inspected and "cleaned" (by human beings, not machines). The "cleaning" involved such activities as erasing stray marks on answer sheets, reconstructing answer sheets that had somehow been damaged in transit, and transferring responses to intact blank sheets if necessary; also making certain that the necessary identification information (school code, etc.) was on each answer sheets before it was separated from the identifying materials with which it was packed. The arrangements called for all this to be done under AIR auspices, rather than by MRC. Therefore Marion Shaycoft spent a month in Iowa City in the spring of 1960, hiring and training a large crew of part-time clerks (mostly housewives who wanted to earn extra money) to do this processing. She had enough faith in the TALENT tests that she used an appropriate subset of them in selecting the personnel. This worked very well. In due course she appointed as the supervisor of this operation one of the clerks she had hired who had proven especially competent and who was able to work full-time. That was Marjorie Major. The operation she was to supervise came to be known informally as the "TALENT Laundry" because of the answer-sheet clean-up activities involved. It was a store-front operation, arrangements having been made to use for the purpose a vacant store a couple of blocks away from the University of Iowa building in which MRC's machines were located. The TALENT Laundry, under Marjorie Major's capable supervision, was to operate for the better part of a year.

In addition to operations of the answer-sheet-inspection-and-cleanup type that MRC required as a preliminary to the machine passes, one very important activity that the "Laundry" staff performed was to riffle through each pack of A and C answer sheets and pull out the ones for which the last two digits of the six-digit student identification number were 00, 25, 50, or 75. This provided a 4% sample. These answer sheets (the 4% sample) were to be the first ones processed on the scoring machine. Arrangements to give them priority would provide data on which interim percentile norms could be based, for use in reporting results to the schools, on 37 variables.

The remaining answer sheets of each kind were arranged in 11 groups--the 10 school subsamples into which the probability sample was divided (as described in Chapter 2, Section A) and the nonprobability cases--since the plan was that after the processing of the 4% sample A and C answer sheets, subsequent scoring machine processing would proceed in subsample order. Document Reader processing would have to await completion of the Document Reader--i.e., getting the remaining bugs out of it.

2. At MRC

Each of the answer sheets to be processed on the Scoring Machine (Answer Sheets A and C) required two passes to yield the required 52 scores. Answer Sheets B1 and B2 each required three passes through the Document Reader, with double-punched cards as output. Record Form Z only required a single pass.

Thus we wound up with 11 IBM cards per student, or about five million cards! MRC scored the 4% sample as soon as it was provided, and also produced distributions in printout form, so that percentile conversion tables could be computed by grade. After Project staff had prepared the conversion tables, MRC scored the rest of the answer sheets, and prepared printout reports to send to the schools showing each student's raw score on 37 variables, the corresponding interim national percentile for the grade population, and the school mean for each grade.

3. After MRC

When the five million cards were put on tape on the University of Pittsburgh computer (where, by this time, the IBM-650 had been replaced by the IBM-7070, which was far more advanced and had far greater capacity), Glenn Roudabush (the second programmer to have joined the Project staff) found that the easiest procedure was to make two tapes for each double-punched card (with the left half of the card, plus identifying information, on one tape, and the right half, plus the same identifying information, on the other half). Thus the five original documents (answer sheets), which had been inflated to 11 IBM cards, resulted eventually in 17 tape records.

Making and merging these 17 tape records for almost half a million students was an exceedingly complex and lengthy job (it lasted years, as a matter of fact), and it was greatly complicated by the fact that most of the project staff was in Washington, while the computer and programmer were 250 miles away, in Pittsburgh. Further complicating the operation was the fact that it had to take place in a milieu over which Project TALENT had no control--the University of Pittsburgh's Cathedral of Learning, which was where the Computation and Data Processing Center was located. Glenn Roudabush, on whom the burden of the computer work to plan and make the tape master file of 1960 data devolved, found he had to work nights, for months on end, because that was the only way he could get adequate access to the computer.

Meanwhile all kinds of logistical problems were interfering with speedy completion of the task. One problem was that a single operation couldn't be carried through on all cases at once (which would have been maximally efficient) because it was necessary to give top priority to completion of one of the 10 probability sample subsamples ("subsample G") so that we would have at least one subset of data in fairly complete form, on which preliminary or exploratory work could be done. Another problem resulted from the lack of convenient space in which to store the IBM cards and keep processed and unprocessed ones separate. For instance on one memorable occasion after Glenn thought he had finished a particular sequence of computer runs, several boxes of unprocessed IBM cards from MRC were discovered in a broom closet! (He had to start the sequence all over again.)

To find out what happened next in this long-running soap opera, see Chapter 13!

Chapter 8. WITH A CAST OF THOUSANDS!

This chapter is intended to function as a guide to the cast of characters--the project staff and some of the others who played a role in the activities and events with which this History is concerned. The chapter title is perhaps only a slight exaggeration, even though somewhat under a thousand names are mentioned.

A. The staff

Figure 8-1 provides in tabular form an overview of the professional, technical, and supervisory staff. The titles shown in this figure are intended primarily as functional indicators. They are not necessarily identical in all cases with the actual officially assigned titles, since the official titles used have tended to change from time to time. For instance officially Dick Williams was the "Director of Computer Operations" while Gary Fulscher was "Manager of Data Processing". But it was essentially the same position, so both are shown under the "Director of Computer Operations" heading.

Figure 8-2 is a time-and-position framework, which gives some idea of what the composition of the staff was in any given period, at least with respect to the positions listed.

Figure 8-1. Project Professional, Technical, and Supervisory Personnel: 1957-76

PRINCIPAL INVESTIGATOR: JOHN C. FLANAGAN

Early phase: Instrument development, sample selection, test administration, one-year follow-up

Institute for Research in Education Director 1963-64: Frederick B. Davis* — Project Director 1958-64: John T. Dailey

PRE-PROJECT PRELIMINARY PLANNING	INSTRUMENT DEVELOPMENT	EXPERIMENTAL TRYOUT	SAMPLE SELECTION	TEST ADMINISTRATION	SENIOR RESEARCH STAFF	DATA PROCESSING AND ANALYSIS: PRE-COMPUTER PROCESSING	DATA PROCESSING AND ANALYSIS: COMPUTER PROGRAMMER: DATA PROCESSING
Project Director: Robert C. Craig 1957-58	Test Development: Marion F. Shaycoft Aptitude and Achievement Test Editor Dorothy S. Edwards Mary B. Willis Student Questionnaire Development: David B. Orr Student Activity Inventory Editor Isadore Goldberg Interest Inventory and Student Information Blank Editor School Questionnaire Development: David Votaw, Sr. Consultant David B. Orr General School Characteristics Questionnaire Editor ~~Dorothy S. Edwards~~ Isadore Goldberg Guidance Program Questionnaire Editor	Test Administration Supervisor: William A. Gorham Data Analysis Director: Marion F. Shaycoft	Computer Programmer: Gary Lotto	Supervisor: William A. Gorham	Director of Measurement Research: Marion F. Shaycoft Special Studies: William A. Gorham School Characteristics Studies: Isadore Goldberg David B. Orr Manpower and Guidance Studies: Isadore Goldberg Supervisor for Services: Clinton A. Neyman, Jr.	Marion F. Shaycoft Director Marjorie Major Supervisor	Glenn E. Roudabush
					JUNIOR RESEARCH, EDITORIAL, AND SUPERVISORY STAFF Stuart E. Sherman Dagny K. Henderson Hedwig Oswald Ida G. Kator Richard A. Williams		COMPUTER PROGRAMMER: DATA ANALYSIS Gary Lotto Glenn E. Roudabush George R. Burket

* On sabbatical from Hunter College

Continued on next page

Figure 8-1 (continued)

PRINCIPAL INVESTIGATOR: JOHN C. FLANAGAN

Subsequent phases: Five-year and eleven-year follow-ups, and planning for subsequent stages

PROJECT DIRECTORS

William W. Cooley
1964-67

Abraham Carp
1967-68

John C. Flanagan
1968-71*

David V. Tiedeman
1971-73

William V. Clemans
1974-76**

*During this three-year period, John C. Flanagan was Acting Director as well as Principal Investigator

ASSOCIATE DIRECTORS

Marion F. Shaycoft#
1965-76*

James M. Richards, Jr.
1969-71

Donald H. McLaughlin#
1974-76

Wendy M. Yen#
1974

Sandra R. Wilson
1974-75

Lauress L. Wise#
1975-75**

*Director of Measurement Research 1960-65

DATA BANK DIRECTORS

Richard W. Holdeman
1964-65

Lyle F. Schoenfeldt
1965-68

John G. Claudy
1969-75

Donald H. McLaughlin##
1975-76**

ASSOCIATE DATA BANK DIRECTORS

Mark M. Greene
1967-68

Marsha D. Brown###
1968-69

OTHER SENIOR RESEARCHERS

Louise W. Cureton
1963-76

Paul P. Mok
1964-65

Charles E. Hall
1965-67

Paul R. Lohnes
1965-67

Alan E. Bayer
1967-68

Michael Wargo
1971-72

G. Kasten Tallmadge
1972-73

Donald P. Horst
1972-73

Robert J. Rossi
1974-76

Kevin J. Gilmartin
1975-76

SENIOR EDITORIAL SERVICES

Mary B. Willis
1971-74

MISCELLANEOUS ADMINISTRATIVE SERVICES

Kenneth W. Buck
1975

COMPUTER PROGRAMMING***

Data Analysis Director:

Bary G. Wingersky
1965-67

Data Processing Director:

Jon vanWormer
1964-66

Marilyn S. Wingersky, 1966-67###

General Supervision of Computer Operations:

James M. Richards, Jr., 1969-70

Director of Computer Operations:

Richard A. Williams, 196[illegible]70###
John G. Claudy, 1970-73
Gary V. Fulscher, 1973-74

Special One-Man Task Forces (Seriatim) for Documentation:

Richard T. Johnson, 1972
Calvin E. Wright, 1972
Donald P. Horst, 1972-73

DATA BANK RESEARCH ASSISTANTS

Pittsburgh:

Mary K. Garee

Palo Alto:

Susan K. Genung
Ardys J. Bloomquist

JUNIOR RESEARCH, EDITORIAL, AND SUPERVISORY STAFF

Pittsburgh:

Janet G. Combs
Carol J. Dalcanton
Judith D. Miller
Deborah Tien

Palo Alto:

Elizabeth A. Reid
Joan M. Altick
Jay Egan
Paulette Doudell
Nancy Carr
Martha Gregory
Kathleen Carrel
Clifford Potts
Patricia Maloney

OTHER RESEARCH ASSISTANTS

Pittsburgh:

Susan J. Becker
Ronya J. Sallade
J. Fred Beamer
Susanne R. Lewis
Robert H. McMillen

Palo Alto:

Constance A. Palmer
George K. Posey
David H. Bain
Mykol C. Hamilton
Rebecca D. Strause
William S. Taylor
Kathleen L. Williams

PROGRAMMERS

Pittsburgh:

Carl Kuehne
Arleen M. Lichtenstein
Seon H. Cho
Francis Wimberly

Palo Alto:

Michael J. Quirk
Michael Barenfeld
Patrick McDonnell
Gary V. Fulscher
Jay A. Woods
Wendy B. Bartlett

JUNIOR PROGRAMMERS, ASSISTANT PROGRAMMERS, AND TRAINEES

Pittsburgh:

Robert L. Hawk, Jr.
James A. Fierst
William G. Grandi
Robert I. Fink
Paul Bianculli
Karl D. Larsen, Jr.
Wesley Rohrer III
Richard D. Heim
Janice B. Dunn
Susan N. Barclay
Marjorie Ginsburg
Robert J. McGrail
Barbara M. Simon
Charles H. Welch

Palo Alto:

Susan J. Freehafer
Karen M. Guckes
Judith L. Miller
Ardys Bloomquist

#	On project research staff earlier, in role other than Associate Director.
##	On project research staff earlier, in other role.
###	On project staff earlier in nonsupervisory capacity.
**	Continuing
***	The computer programing staff is shown up to 1974. In 1974, Project TALENT ceased to have its own programming staff and instead used AIR's Palo Alto computing staff.

Figure 8-2. The Changing Cast: 1957 - 1976

	1957-1958	1958-1960	1960-1964	1964-1965	1965-1966	1966-1967	1967-1968	1968-1969	1969-1970	1970-1971	1971-1973	1973-1974	1974-1975	1975-1976
Principal Investigator	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF	JCF
Project Director	RCC	JTD	JTD	WWC	WWC	WWC	AC	↓	↓	↓	DVT	WVC	WVC	WVC
Associate Director					MFS	MFS	MFS	MFS	MFS JMR	MFS JMR	MFS	MFS	MFS* DHM** SRW** WMY**	DHM# MFS## LLW###
Director of Measurement Research			MFS	MFS										
Test Editor		MFS												
Associate Test Editor		DSE												
Data Bank Director				RWH	LFS	LFS	LFS	LFS	JGC	JGC	JGC	JGC	JGC	DHM
Associate Data Bank Director							MMG	MDB						
Computer Operations Director		GL		JvW			RAW	RAW	RAW	JGC	JGC	GVF	***	***
For Data Analysis			GL GRB		BGW	BGW								
For Data Processing			GER		JvW	MSW								
Tape improvement and documentation task forces											RTJ DPH			MFS#### LLW

---

* Initially

** Later

*** Not a Project TALENT staff member

# For Data Bank

## For data tape improvement and documentation

### Other

#### Initial planning

KEY

MDB	Marsha D. Brown
GRB	George R. Burket
AC	Abraham (Al) Carp
JGC	John G. Claudy
WVC	William V. Clemans
WWC	William W. Cooley
RCC	Robert C. Craig
JTD	John T. Dailey
DSE	Dorothy S. (Jo) Edwards
JCF	John C. Flanagan
GVF	Gary V. Fulscher
MMG	Mark M. Greene
RWH	Richard W. Holdeman
DPH	Donald P. Horst
RTJ	Richard T. Johnson
GL	Gary Lotto
DHM	Donald H. McLaughlin
JMR	James M. Richards, Jr.
GER	Glenn E. Roudabush
LFS	Lyle F. Schoenfeldt
MFS	Marion F. Shaycoft
DVT	David V. Tiedeman
JvW	Jon vanWormer
RAW	Richard A. Williams
SRW	Sandra R. Wilson
BGW	Bary G. Wingersky
MSW	Marilyn S. Wingersky
LLW	Lauress L. Wise
WMY	Wendy M. Yen

1. The Principal Investigator

John Flanagan, as the person who conceived and planned the project, and as the prime mover behind it, has naturally been the Principal Investigator throughout the nearly 20 years that have elapsed (as of the end of 1976) since the Project started. The continuity that he has provided has kept the Project on course all these years. There have been some small shifts in plans, some minor changes in details along the way, but the general features and the goals have remained substantially the same. Without the constant element provided by having the same Principal Investigator from the very beginning this would have been difficult, if not impossible, to achieve.

2. Project Directors

As has already been indicated, Robert Craig was the first Project Director. During his single year in that role his activities were largely exploratory--investigating what sorts of procedures might be feasible and what sorts of tests and questionnaires might be useful.

He resigned in 1958 to accept an appointment at Marquette University. (He is now at Michigan State.)

John Dailey, the next Project Director, had had 16 years of experience in the Federal government. During World War II he had been a psychologist in the Aviation Psychology Program headed by John Flanagan. After the war, most of the psychologists left; he stayed at Lackland Air Force Base until 1951, directing the Personnel Research Unit. After that he had seven years in Washington, with the Bureau of Naval Personnel (1951-58), where he was technical director of the Research Division. He gave up Civil Service tenure to accept the appointment as Project Director of the project, still unfunded beyond the planning stage, and with no promises from any funding agency. John recognized that he was taking a big gamble, but the prospective project fitted in exactly with his professional interests, and directing it was, he says, what he most wanted to do. He did not regret his decision. Around 1947, when he was at Lackland, he had been campaigning for a large-scale testing of a national sample. He had recognized the desirability of having data based on a general population, not just on those who happened to be in the Air Force. The project that he was leaving the Bureau of Naval Personnel to direct was exactly in line with that idea. Furthermore he had been one of the participants in the initial planning conference for the project, in February 1957. He had been excited by the idea unveiled by John Flanagan at that conference. These factors combined to make him enthusiastic about the project and the opportunity to direct it. And during his six years as director of Project TALENT that enthusiasm never wavered.

When the project moved to Pittsburgh, however, John decided his ties in the Washington area were too strong for him to be interested in making the move. Consequently he resigned from AIR (and the University of Pittsburgh). He had been TALENT's Project Director for a much longer period than anyone else before or since.

John accepted an appointment at George Washington University as Research Professor of Education, and Director of the Education Research Program. (In that capacity he carried out a number of studies that were based on Project TALENT data, and utilized the services of the Data Bank.) After four years at GW he became a psychologist for the Federal Aviation Administration, where he is now the Chief of the Analytic Branch of the Behavioral Sciences Division of the Office of Aviation Medicine. He has done some interesting work there, including developing a behavioral profile of potential skyjackers (or as the FAA prefers to call them, hijackers) which was instrumental in reducing the air piracy threat, at airports. (In 1972 he received the Flight Safety Foundation's top safety award for this work.) He is now working on the somewhat more general problem of protection against terrorist activity, and is also developing new types of tests for air traffic controllers.

Perhaps this would be a good place to mention that one of John Dailey's major contributions to Project TALENT was the concept of "follow-back studies" that he worked out. (This concept is discussed briefly in Chapter 21 and also in Chapter 13. It has not been used as extensively or as effectively as it should have been--but that is the fault of neither the concept nor its originator.)

Before going on to the next Project Director, we must say something about the important role that Fred Davis played. During the academic year 1963-64, Fred was on sabbatical from Hunter College. During that same year, members of the TALENT staff were scheduled to write a very large research report (it turned out to be over 700 8½" x 11" single-spaced pages) entitled "The American High School Student". Fred, who was a first-rate psychometrician, test-writer, and educational researcher, was also an expert editor. As a matter of fact, he had been the editor-in-chief of the entire

19-volume series of Army Air Forces Aviation Psychology Program Research Reports produced after the end of World War II. Fred was also a long-time member of AIR's Board of Directors. The coincidence of Fred's sabbatical coming at the same time as the enormous report-writing task that lay ahead of the project staff was a happy one for Project TALENT. John Flanagan persuaded Fred to spend his sabbatical year in Washington, as editor-in-chief of the forthcoming report and as director of an Institute for Educational Research (IRE), which didn't yet exist but would be created as one of the American Institutes for Research's several institutes. Project TALENT, along with some related activities, would be put under the umbrella of the Institute for Educational Research.

Thus during the last year in Washington (1963-64), Project TALENT had a tripartite management structure: Principal Investigator, Institute Director, Project Director. The arrangement worked very well; however after the project moved away from Washington the IRE Institute Director and the Project TALENT Project Director were invariably the same person.

We hoped that Fred would decide to stay with Project TALENT when his sabbatical ended, instead of going back to Hunter. As things turned out, however, he decided not to go back to Hunter, but also to withstand the heavy persuasion efforts to get him to stay at AIR. Instead, he accepted an appointment to the faculty of the University of Pennsylvania. After he left AIR, Fred maintained his interest in Project TALENT, and was a consultant to it on many occasions. He died in the spring of 1975, and we still feel the loss.

After John Dailey, the next Project Director was William W. Cooley, who had been teaching at Harvard. He came to AIR in 1964.

Bill's main professional interest was multivariate analysis. He and Paul Lohnes had written a book on multivariate analysis in which not only the statistical techniques but also computer programs they had prepared for implementing those techniques were presented. (More about those programs later; see Chapter 14.) For the three years that Bill Cooley was Project Director, most of the research done on TALENT data utilized multivariate methodology. When Project TALENT was about to move to California, Bill Cooley, like his predecessor, decided he would rather stay where he was. He preferred Pittsburgh to Palo Alto*. Therefore he resigned from AIR in 1967, to devote full time to his professorship at the University of Pittsburgh. (He had been on the faculty there since 1964, but directing Project TALENT had been his main occupational activity.) He is now Co-Director of the University of Pittsburgh's Learning Research and Development Center.

When the Project moved to Palo Alto, in the summer of 1967, Abraham (Al) Carp became the new Project Director. Like John Dailey's, Al's professional experience was heavily Air Force. He had been at AFPTRC since 1951 (the year John Dailey left) in varying capacities, becoming in 1957 the technical director of the Personnel Laboratory, Aeronautical Systems Division.

Al was Project Director for only a year. In the fall of 1968 he became the Director of Research in ETS's Berkeley Office.

There was a three-year hiatus (1968-71) before his successor was appointed. During the interim, John Flanagan functioned not only as Principal Investigator but also as Acting Project Director.

In the summer of 1971 David Tiedeman resigned from a professorship at Harvard, to become Project TALENT's Project Director.

---

* Californians will not understand!

One of Dave's principal activities while he was at AIR was to work on the preparation of the Career Data Book, a large volume designed to present Project TALENT's 5-year follow-up data for career groups in a manner that would facilitate its use in career guidance. Dave devoted a lot of effort to the Career Data Book although it was far from finished with he left AIR. (Dave also devoted considerable attention to getting a "Systems Manual" developed for use by IRE and specifically for use in Project TALENT operations. (The Systems Manual was written almost wholly by Calvin Wright, who had already been on the AIR staff for about 5 years when he joined Project TALENT for a very brief period.) Dave stayed at AIR for two years, leaving in the summer of 1973 to accept an appointment as professor at the University of Northern Illinois.

Several months later, William V. Clemans accepted an appointment as the next Project Director. Bill, a vice-president of Science Research Associates (a Chicago publisher of tests and educational materials), was no stranger to AIR. In the early 1950's he had worked for a short time on an AIR project that took him to Korea for several months. And in the late 1950's and early 1960's, when AIR developed some examinations for the National Board of Medical Examiners, Bill Clemans, who was then the Director of the Testing Service for that organization, monitored the project.

Like most of his predecessors Bill has a strong background in psychometrics, having received his doctorate at the University of Washington, where he was one of Paul Horst's students.

Unlike some of his predecessors, such as John Dailey and Bill Cooley, both of whom were heavily involved in doing research on Project TALENT data (as, too, the Principal Investigator has been)

Bill Clemans has regarded the job of Project Director as primarily administrative in nature. (This difference is mentioned only to point up the not too surprising fact that different project directors may have very different operating styles.)

Bill is the current Project Director.

3. Associate Directors

Marion Shaycoft, who became an Associate Director of Project TALENT (and of IRE) in 1965, had joined the project staff in 1958. She was Editor-in-Chief during the test development (i.e., until sometime in 1960) and then she became the project's Director of Measurement Research. Thus when she became Associate Director she was thoroughly familiar with the salient aspects of the project.

In 1969 it became advisable to appoint a second Associate Director, to take over one of the functions that the Project Director had previously performed--namely general supervision of the computer operations. This responsibility devolved upon James M. (Mac) Richards, a new member of the research staff. The Director of Computer Operations reported to him. Mac performed this Associate Director function for only about a year however. In 1970 other responsibilities--specifically direction of a new project that had nothing to do with Project TALENT--required him to cut back sharply on the amount of time he could devote to Project TALENT. Mac left AIR in 1971, to accept a position at the University of Missouri at Kansas City. Again the Director of Computer Operations reported directly to the Project Director. Thus the project was down to one associate director again. But in 1974 there was a sharp inflation in the number. During the period from 1974 through 1976 there were usually three associate directors, and occasionally four.

Donald McLaughlin, who had been teaching at UC Berkeley for several years, was one of the new Associate Directors; Wendy Yen, a new Ph.D. from Berkeley, who had been on the TALENT staff about half a year, was another. She left after about half a year as Associate Director, to join the staff of CTB/McGraw-Hill (test publishers, in Monterey, California). Don McLaughlin, fortunately, is still associated with the project (primarily as Director of the Data Bank). Sandra Wilson, who had been on the AIR staff for several years but had not been associated with Project TALENT, transferred to the Project as another Associate Director, around the middle of 1974. (Project TALENT was the loser when Sandy left the project staff about a year later, to play a major role in AIR's development of new tests for medical school admission.) About that time Lauress L. Wise was added to the lengthening list of associate directors. As discussed later in this chapter, in the spring of 1976 Marion Shaycoft almost completely discontinued active participation in the project's ongoing activities, to concentrate on other activities, including among other things the writing of a monograph on criterion-referenced tests (Shaycoft, 1976) and the writing of this History. In regard to the latter, possibly standing back from the thick of things improves one's perspective as the observer and historian of an activity; at least we hope that that is the most potent effect (although some impairment of the view is probably an inevitable byproduct of the distance).

4. Data Bank Directors

A few projects of the "Data Bank type"--projects in which computer output was provided at cost for some outside researcher (i.e., not an AIR staff member) who wished to use TALENT data--were done while Project TALENT was still based in Washington. These projects were handled primarily by George Burket, one of our computer programmers. But it was not until the move back to Pittsburgh in 1964 that the Data Bank as such was formally organized*. Its first director was Dick Holdeman, who was then a new Ph.D. from Purdue. He stayed about a year, then resigned to accept a job as an industrial psychologist. His successor was another new Ph.D. from Purdue, Lyle Schoenfeldt. Lyle stayed until about the end of 1968--which means that he was one of the staff members who made the move to Palo Alto in 1967 (see Chapter 9). He left to accept a position on the faculty of the University of Georgia. (He is now at Rensselaer Polytechnic Institute.) When Lyle directed the Data Bank, Mark Greene was Associate Data Bank Director for a year (1967-68) at the start of the Palo Alto phase, and Marsha Brown, who had started at the same time as Mark, was Associate Director of the Data Bank the following year (1968-69). Mark left in the fall of 1968, to join the staff of Northwest Regional Laboratory. Marsha left the following year, to go back to school for her doctorate. She was at the University of Wisconsin for a year, but got her doctorate at Harvard, where she worked with Christopher Jencks.) Lyle Schoenfeldt's successor as Director of the Project TALENT Data Bank was John Claudy, who was a new Ph.D. from the University of Tennessee. (John had been one of the students of Ted Cureton, who had been associated with Project TALENT as a consultant and had done some special studies involving TALENT data, in Knoxville.) John Claudy is still at AIR, but he transferred to other activities and discontinued his functions as Data Bank Director in 1975, when Don McLaughlin took over in that capacity. Don is the current Data Bank director.

5. Other recent senior staff members

Mary Willis (the same Mary Willis who was mentioned in Chapter 5 as having drafted the Mechanical Reasoning items, and in Chapter 6 as having written the two "domain tests" of reading) was a regular full-time member of the Project TALENT staff for three years (1971-74) after which she transferred to another AIR project again. During the three years Mary was on the Project TALENT staff she had editorial responsibilities of a varied nature. She was the editor of the annual edition of the Project TALENT

---

* See Chapter 22 for a discussion of the Data Bank.

News (a four-page "newspaper" sent to all project participants), and she wrote most of it. She also edited a Project TALENT Bulletin, and was heavily involved in the writing of the Career Data Book. Mary is a highly skilled writer. (She was a coauthor of a book entitled "Secret Archives of the Vatican" which, when it came out, was a Book-of-the-Month Club alternate selection.)

Kenneth Buck, who has a doctorate from Claremont Graduate School, and whose major fields were political science and education, was on the staff for about a year, starting in 1975. His role was as Senior Management Specialist; his duties were largely administrative.

6. Personnel in charge of computer operations

When the project was located in Washington except for the computer in Pittsburgh, we had three programmers--Gary Lotto, Glenn Roudabush, and George Burket. Glenn was primarily responsible for the data-processing tasks involved in preparing master tapes containing the raw data, while Gary and George were more involved in other activities, such as data analysis. (The distinction between "data processing" and "data analysis" may be a slight one, but it is convenient nonetheless. It was particularly convenient in Pittsburgh, because the University had two entirely different computers--the IBM-7070, which was used primarily for data processing, and the IBM-7090, which was used for data analysis.)

Gary, the first of the three to join the staff, was also the first to leave. (He left in 1963, to accept a position with IBM.) George Burket left in 1963 to accept a position in the test division of Science Research Associates, where he joined the staff of Bill Clemans, the present Project Director. In 1964, about the time that the project moved from Washington to Pittsburgh, Glenn Roudabush, too, resigned. He joined the staff of California Test Bureau (now

CTB/McGraw-Hill) where he remains. (George Burket has been there, too, since about 1967.)

Thus, except for a couple of programmer trainees there was no carryover in computer staff when the project moved to Pittsburgh.

Jon van Wormer was the first programmer hired; Bary Wingersky was hired shortly thereafter. Their responsibilities were split down the middle, with Jon being responsible for the data processing and Bary for data analysis. Bary, who is an extremely talented programmer, and the staff working under his direction, wrote all the programs that the research staff needed, including some very complex sequences of multivariate analysis programs. The programs were fast, accurate, and efficient; they worked beautifully. Ready-made packages (such as BMD) were never used while he was on the staff. In effect he developed his own package designed especially to meet the needs of the TALENT staff.

When Jon van Wormer left in 1966, Marilyn Wingersky (Bary's wife and a skilled programmer in her own right) was promoted to fill the vacancy. (Marilyn had been on TALENT's programming staff for about a year, so she was thoroughly familiar with the situation.)

When the project moved to Palo Alto in 1967, Bary and Marilyn Wingersky decided they preferred to live in the East, so they both accepted job offers from Educational Testing Service, in Princeton.

Meanwhile Dick Williams, one of the two staff members who had moved with the project from Washington to Pittsburgh*, had been learning programming. In Washington he had been an administrative assistant and then a research assistant; in Pittsburgh he had started out as a computer trainee; by the time his three years in

---

* The author of this History was the other one.

Pittsburgh were completed he was a full-fledged programmer. Thus when the move to Palo Alto left the project without a programming supervisor, Al Carp promoted Dick to that position. It was a logical decision; Dick was clearly more familiar with the TALENT tapes and TALENT procedures than any newcomer could have been. When Dick left in 1970 to accept a job as programmer for a pharmaceutical company John Claudy was his successor as Director of Computer Operations. John was already Director of the Data Bank, and he continued in that role, too. He continued to wear the "Director of Computer Operations" hat until 1973, and the Data Bank hat until 1975. In 1973 Gary Fulscher, a senior programmer on the staff, who was working for his doctorate at Stanford, was promoted to Director of Computer Operations. In 1974, by an administrative decision, the entire programming staff was made independent of Project TALENT, and that is why this particular section of the chapter ends at this point!

7. Computer trouble-shooter task forces

Now we come to the point where we should say something about our series of one-man task forces, in Palo Alto, for the solution of our tape documentation and other computer-related problems. As the reader will learn in Chapter 13, tape format documentation, records, and allied tasks associated with the management of computer operations were, it seemed, never-ending problems. Therefore at verious times one-man task forces were appointed to make what progress they could towards amelioration of the trouble. Rick Johnson, a psychologist who was knowledgeable about computers and had been on the AIR staff for several years (though not on the TALENT staff) was the first appointee to this difficult assignment. And he made real progress on it, by producing the best tape inventory we had ever had (some might say the only one we had ever had). This task required hard work, carefulness, and accuracy--

all of which Rick provided. He also began work on a Standards Manual. When he left Project TALENT in 1972, Calvin Wright stepped into the breach; his assignment was to complete the Standards Manual that Rick had started. This was to be useful in maintaining appropriate standards of accuracy and thoroughness in programming, checking, and documenting, for all computer-related operations. As was touched upon earlier, Cal converted the embryo "Standard Manual" into a massive "Systems Manual", which was not limited in its scope to computer-related operations. The third one-man task force for TALENT's computer operation was Don Horst, who was assigned this role when he returned in 1972 from an AIR assignment in the Far East (mostly in Singapore). Don, unfortunately, did not have much time to devote to operation of the computer operation, since his other major assignment from the Project Director, which was to implement parts of the Systems Manual (mostly the non-computer-related parts) throughout Project TALENT, proved to be inordinately time-consuming--through no fault of Don's, needless to say, since his actions were fully prescribed by the Systems Manual. Don left Project TALENT, and AIR, in 1973.

In 1975 we received a generous grant from NIE explicitly for improving the tapes and their documentation, and carrying out certain related tasks such as the preparation of an improved handbook for Data Bank users. The grant was a portion of a considerably larger grant, and it was earmarked for this "tape clean-up" activity.

Marion Shaycoft, who had written the proposal for this activity and had campaigned for it (or for something like it) for years, was the Director. Her staff consisted of Lauress Wise and Wendy Bartlett. Wendy, who had been Gary Fulscher's successor as the Director of Computer Operations (after this ceased to be a TALENT function), had recently resigned to go back to school, but she was available part-time as a consultant. Laurie Wise is a skilled programmer (as well as being one of Project TALENT's proliferating associate project directors).

Readers who have a morbid curiosity about some of the computer-related problems that have been touched on in this section are invited to read Chapter 13.

8. Other research staff members

a. Washington phase

The roles of Bill Gorham, Jo Edwards, Izzy Goldberg, and Dave Orr, who were among the earliest additions to the TALENT staff after John Dailey became Project Director have already been mentioned--Bill Gorham in Chapter 7 and the other three in Chapter 5. Of the four, Jo Edwards is the only one who is still at AIR.

Bill Gorham left AIR to accept a job with the U.S. Civil Service Commission, where he is now Director of the Personnel Research and Development Center. Izzy Goldberg is Professor of Community Education at Federal City College, in Washington. Dave Orr, after a number of other activities, is now at NCES, where the population data from which the TALENT sample was drawn came from.

Two other members of the professional staff should be mentioned. Clinton A. (Ted) Neyman, a graduate of the

U.S. Naval Academy, had retired as a captain from his first career, and chose educational research as his second. Actually, though, when he joined Project TALENT in 1960 as the Supervisor for Services (a largely administrative role) the field of psychometric research was by no means new to him. Captain Neyman, before his retirement from the Navy, had been assigned for many years to the Bureau of Naval Personnel, where he directed a section responsible for psychometric research. And under Navy aegis, he had been sent to Stanford University, where he had earned an M.A. in 1949 in educational research. Ted stayed at AIR until 1964, gradually adding research activities to his administrative assignments. He was a valuable staff member. When Project TALENT moved to Pittsburgh, Ted remained in Washington; he moved with John Dailey from AIR to George Washington University, where he was on the staff of the Educational Research Program. Simultaneously he worked towards his doctorate at GW. He received his Ed. D. there in 1971, and went on to direct the Educational Research Program at GW after John Dailey left. It is sad to have to report that after this rather impressive mid-life career-change level of achievement, he died, much too young, a few years later.

Another staff member from the project's Washington period was Dan Sillers, a counseling psychologist. Dan doesn't show up in Table 8-1, because he was associated with the project so briefly. The only reason he is mentioned here is that to the best of my knowledge he is the only former staff member who became a college president. After leaving Project TALENT, he joined the faculty of Jamestown College in North Dakota, and was appointed President a couple of years later.

b. Second Pittsburgh phase

Some members of the research staff when the project was in Pittsburgh have not yet been mentioned and should be. Paul Lohnes, who was Bill Cooley's coauthor on their book on multivariate analysis (Cooley and Lohnes, 1962) joined the project in 1965. Like Bill, Paul had a teaching appointment at Pitt. Like Bill, Paul was (and probably still is) a multivariate-analysis enthusiast. Like Bill, when the time came to move to Palo Alto, Paul decided he didn't want to leave the East (although he was willing to move away from Pittsburgh). He decided to rejoin the faculty of the State University of New York at Buffalo, where he had been immediately before joining the TALENT staff. His reason for preferring New York State to California wasn't just that he considered the Atlantic a better ocean than the Pacific. Probably no small consideration was the fact that at SUNY he would have more opportunity to teach than at Pitt--and he gives every indication of really enjoying teaching.

Charles (Chuck) Hall was also on the project staff from 1965 to 1967. After he left AIR he was with ETS in Princeton for a while. He is now with Abt Associates, to the best of my present knowledge.

Paul Mok was one of Project TALENT's more memorable staff members. Before joining the Project TALENT staff, Paul had been "chief of party" in a one-person "party" on an AIR project in Liberia, where he and his family spent a couple of years. Although he was only on Project TALENT about four months (late fall of 1964 to early spring of 1965) he left his mark on the project. A prolific writer, Paul had written several popular books about education (including "A View from Within" (1962) and "Pushbutton Parents and the Schools" (1964)) before joining the AIR staff. When he left Project TALENT, it was to take a position in New York with a firm of consulting psychologists and to write a novel. If Dan Sillers is Project TALENT's only alumnus who became a college president, Paul Mok is probably its only published novelist.

c. Palo Alto phase

For about a year, Project TALENT had a resident sociologist on the staff: Alan E. Bayer. (He was already familiar with the project because he had been a Data Bank client previously.) While on the project staff, he used TALENT data for some interesting research on marriage and some research on college dropouts. When he left in 1968, it was to join the faculty of Florida State University.

Mike Wargo, who had been on the AIR staff for a couple of years and had done an excellent (though controversial) reanalysis and synthesis of compensatory education evaluation data (Wargo, Tallmadge, Michaels, Lipe, and Morris, 1972), joined the TALENT staff in 1971. His chief assignment was to develop a tentative plan for a study of the careers of scientists, since NSF had provided a grant for that purpose. When Mike left in 1972 to take a position with the Office of Education, Kast Tallmadge, who had been at AIR since 1966 and had done lots of interesting and worthwhile studies involving, among other things, evaluation of compensatory education, the development of training procedures, exploration of aptitude-treatment interactions, and instructional technology, joined the TALENT staff and took over on what Mike Wargo had been doing. (Kast left, in 1973, to establish and direct a Western office for RMC Research Corporation. Don Horst left at about the same time, to join him in that venture. Marion

Shaycoft inherited the NSF study from Kast, completed the planning, and completed the study*.)

Two other fairly recent additions to the Project TALENT staff are Bob Rossi and Kevin Gilmartin, both of whom received their doctorates in 1974. Bob's, from Stanford, is in philosophy of education (making him our first genuine certified philosopher!) while Kevin's, from Carnegie-Mellon, is in cognitive psychology--which, at Carnegie-Mellon, includes a heavy dose of computer modeling, so that Kevin may be the only one at AIR who knows much about that topic.

d. Washington, Pittsburgh, and Palo Alto

Louise Cureton has been on the staff since 1963, and even before that she was associated with Project TALENT as a Regional Coordinator (in Knoxville). Thus it is impossible to classify her as belonging in the Washington, Pittsburgh, or Palo Alto phase; she belongs in all three. Throughout all three phases, however, Louise has remained in Knoxville, where she directs our small but important "Knoxville detachment". Louise has carried out several studies based on the Knoxville data, supplemented by special follow-up data which she collected herself. Among the studies she has done are the following:

1) A study of the early identification of behavior problems
2) A study of welfare dependence
3) A study of the effects of having alcoholic parents

We hope she will continue doing these interesting and worthwhile studies, and that she will continue to take advantage of that valuable research resource, the TALENT data on the Knoxville and Knox County population, that she has done so much to *make* a valuable research resource.

---

* Shaycoft (1975)

B. Early advisory panels

The counsel provided by our various advisory panels was crucial to the success of the project.

In Chapter 2 (in Figure 2-1) we listed the members of the four initial advisory panels--Guidance and Counseling Panel, Educational Research Panel, Testing Problems Panel, and Manpower and Sociology Panel--and (in Figure 2-3) the members of the Advisory Council. In Chapter 3 we listed our panel of advisors on sampling. Needless to say, we are very grateful to everyone in these groups; they helped us get the project started.

C. Regional Coordinators

As is explained in Chapter 7 (Section B) in making arrangements with the schools selected for our sample, and in getting the tests administered, we were, to a very great extent, dependent on our Regional Coordinators. The initial group is listed in Figure 8-3. (Over the years, we have continued to count on Regional Coordinators' assistance, in conducting our follow-ups. The list has changed gradually, and our gratitude extends to the newer Regional Coordinators as well as to the group listed.

D. OE and NIE personnel

Until the end of 1971, the chief financial support of Project TALENT came from the Office of Education; after that, the National Institute of Education was established, and from then on NIE has provided the support for the project. Obviously there had to be many people in both organizations who provided help, backing, and moral support, when help, backing, and moral support were what we needed. The project might not ever have gotten started if it hadn't been for the support of the Commissioner of Education, Dr. Lawrence Derthick, in 1958, and the Assistant Commissioner, Dr. Roy Hall, who was also

Figure 8-3. REGIONAL COORDINATORS IN 1960

- BAMA
  - Paul R. Givens, Birmingham-Southern College
  - Virginia Zachert, Sturm-O'Brien, Inc., Auburn
- ARIZONA
  - Richard E. Schutz, Arizona State University, Tempe
- ARKANSAS
  - Carter Short, University of Arkansas, Fayetteville
  - Hardy C. Wilcoxon, University of Arkansas, Fayetteville
- CALIFORNIA
  - Alex D. Aloia, Loyola University of Los Angeles
  - John Caffrey, Director of Research, Palo Alto Unified School District
  - Frederick J. McDonald, Stanford University
- COLORADO
  - Anthony C. Tucker, University of Denver
- CONNECTICUT
  - Joseph Raymond Gerberich, University of Connecticut, Storrs
- DELAWARE
  - Arthur R. DeLong, University of Delaware, Newark
- FLORIDA
  - Edward Caldwell, Board of Public Instruction, Manatee County, Bradenton
  - Richard H. Kicklighter, University of Florida, Gainesville
  - John V. McQuitty, University of Florida, Gainesville
- GEORGIA
  - Cameron Fincher, Georgia State College for Business Administration, Atlanta
  - G. H. Fort, Board of Education, Atlanta
  - R. T. Osborne, University of Georgia, Athens
- IDAHO
  - Elwyn DeLaurier, State Guidance Supervisor, Department of Education, Boise
- ILLINOIS
  - N. L. Gage, University of Illinois, Champaign
  - Lyman J. Smith, Illinois State Scholarship Commission, Deerfield
- INDIANA
  - N. A. Fattu, University of Indiana, Bloomington
  - H. H. Remmers, Purdue University, Lafayette
- IOWA
  - Arthur Mittman, University of Iowa, Iowa City
  - Gordon J. Rhum, Iowa State Teachers College, Cedar Falls
  - Herbert M. Silvey, Iowa State Teachers College, Cedar Falls
- KANSAS
  - Kenneth E. Anderson, University of Kansas, Lawrence
  - Charles B. Watkins, Guidance and Personnel Service, Kansas State Department of Public Instruction, Topeka
- KENTUCKY
  - Ernest McDaniel, University of Kentucky, Lexington
- LOUISIANA
  - Robert N. Vidulich, Louisiana State University, Baton Rouge
- MAINE
  - David R. Fink, Jr., University of Maine, Orono
- MARYLAND
  - Robert C. Lloyd, Baltimore Public Schools
- MASSACHUSETTS
  - Seth Arsenian, Springfield College
  - George S. Elias, Springfield College
  - Edward Scanlon, Division of Counseling Services, Newton Public Schools, West Newton
- MICHIGAN
  - Claude L. Nemzek, University of Detroit
  - Edwin G. Spacie, Central Michigan University, Mt. Pleasant
  - Buford Stefflre, Michigan State University, East Lansing
  - Frank B. Womer, University of Michigan, Ann Arbor
- MINNESOTA
  - Ralph F. Berdie, Director, Student Counseling Bureau, University of Minnesota
  - Robert J. Keller, Director, University High School, University of Minnesota
- ISSIPPI
  - Roscoe A. Boyer, University of Mississippi, University
  - Russell W. Levanway, Millsaps College, Jackson
- MISSOURI
  - Joseph L. French, University of Missouri, Columbia
  - Robert E. Lefton, Psychological Associates, Clayton
- MONTANA
  - William A. Garrison, Eastern Montana College of Education, Billings
- NEBRASKS
  - Warren Baller, University of Nebraska, Lincoln
- NEW HAMPSHIRE
  - Paul McIntire, University of New Hampshire, Durham
- NEW JERSEY
  - Donald W. Campbell, Board of Education, Newark
- NEW MEXICO
  - Virginia Keehan, Department of Education, Santa Fe
- NEW YORK
  - Warren W. Coxe, Delmar
  - S. David Farr, University of Buffalo
  - Percival M. Symonds, Professor Emeritus, Teachers College, Columbia University
  - Clarence M. Williams, University of Rochester
- NORTH CAROLINA
  - Roy N. Anderson, North Carolina State College, Raleigh
  - Junius A. Davis, University of North Carolina, Greensboro
  - Thomas E. Jeffrey, University of North Carolina, Chapel Hill
  - William D. Perry, University of North Carolina, Chapel Hill
- NORTH DAKOTA
  - Ralph H. Kolstoe, University of North Dakota, Grand Forks
  - Robert E. Larson, State Agricultural College, Fargo
  - Grant M. Norem, State Teachers College, Minot
- OHIO
  - Howard B. Lyman, University of Cincinnati
  - Walter S. Nosal, John Carroll University, Cleveland
  - Ray Wood, Columbus
- OKLAHOMA
  - W. R. Brown, University of Oklahoma, Norman
- OREGON
  - J. Spencer Carlson, University of Oregon, Eugene
- PENNSYLVANIA
  - L. Kathryn Dice, Department of Public Instruction, Harrisburg
  - C. Mauritz Lindvall, University of Pittsburgh
- RHODE ISLAND
  - Frances E. Dunn, Brown University, Providence
- SOUTH CAROLINA
  - R. L. Kalmbach, Columbia Public Schools, Columbia
  - W. C. McCall, University of South Carolina, Columbia
  - Donna S. Young, University of South Carolina, Columbia
- SOUTH DAKOTA
  - V. Gregory Rosemont, Huron College, Huron
- TENNESSEE
  - George E. Copple, Vanderbilt University, Nashville
  - Louise W. Cureton, Knoxville
- TEXAS
  - Robert P. Anderson, Texas Technological College, Lubbock
  - H. Paul Kelley, University of Texas, Austin
  - Saul B. Sells, Texas Christian University, Fort Worth
  - Franklin L. Stovall, University of Houston
  - David F. Votaw, Sr., San Marcos
- UTAH
  - Hyrum M. Smith, Department of Public Instruction, Salt Lake City
- VIRGINIA
  - Richard L. Beard, University of Virginia, Charlottesville
  - Donald J. Herrmann, College of William and Mary, Williamsburg
- WASHINGTON
  - William C. Budd, Western Washington College of Education, Bellingham
  - Glen E. Maier, Eastern Washington College of Education, Cheney
- WEST VIRGINIA
  - Walter Jaracke, University of West Virginia, Morgantown
- WISCONSIN
  - Elden A. Bond, Milwaukee Public Schools
  - Robert C. Craig, Marquette University, Milwaukee
  - Ralph H. Tindall, Milwaukee Public Schools, Milwaukee
- WYOMING
  - R. Duane Andrews, Department of Education, Cheyenne

the Director of OE's Cooperative Research Program. After Roy Hall left, his successors as Director of the Cooperative Research Program continued to provide support and help. Dr. David Clark and Dr. Francis Ianni are among those in this category.

At most stages, however, the person with whom we had closest contact was the project monitor. Over the years there have been a great many monitors, some who filled this role for several years and some for just a few months. As a group they were interested and helpful--and their help was indispensable. In mentioning specific monitors we must start with Alice Scates (of the Office of Education)--not only because she was our first monitor, but also because she was our most frequent monitor, having filled that role at least three separate times, with other people functioning as monitors in the interims; and not only has Alice been our only "recidivist monitor" but she has continued to maintain a lively interest in the project even at other times.

Having named the first monitor, let us now mention the last one who held that position during Project TALENT's Office of Education period--or better yet, let us name the last three: Judith D. Weinstein, Susan Klein, and Laurence G. Goebel. And lest anyone get the idea from the fact that three of the four monitors named have female names that some sort of sex discrimination (against men) was operating, let me assure the readers that all of the numerous monitors between Alice Scates and Judy Weinstein were male--as were all the monitors in the periods between Alice's several hitches in that role.

The list of monitors in TALENT's NIE period is shorter, perhaps because the period itself is shorter (only about 5 years, as compared with OE's 14). When NIE was started, Larry Goebel, TALENT's last monitor under OE, transferred to NIE, and continued as monitor for quite a while. This was important because it helped provide continuity. The next monitor was Garry McDaniels, who was the director of NIE's Teaching Division, in the Basic Skills Group. He continued in a general overseeing role, after he appointed a member of his staff, Jane David, as the next monitor. When Jane left NIE, Andrew C. Porter, who had recently become Acting Director of the Basic Skills Group, inherited responsibility for Project TALENT: he appointed Richard Harbeck as monitor. All of the NIE people named--Larry, Garry, Jane, Andy, Dick--have been interested and helpful--and we needed their interest and their help. Dick Harbeck continues as monitor.

The part of the study represented by NSF's supplementary funds had its own monitor, Dr. Charles H. Dickens of NSF. He and a member of his staff, Naomi Sulkin, have had a continuing interest in Project TALENT.

E. TALENT post-doctoral fellows

In 1966-67, Project TALENT ran a training program for three post-doctoral Fellows--William A. Love, David E. Kapel, and Douglas K. Stewart. The training program is discussed in Chapter 19.

Chapter 9. THE GEOGRAPHY OF THE PROJECT

If by "the geography of Project TALENT" were meant the location of the members of the sample, the correct description would be that they are all over the country. Strangely enough, the correct description is almost the same if by "the geography of Project TALENT" is meant the location of Project headquarters. The physical location of project headquarters has changed almost as many times in the years since the project started as has the identity of the Project Director.

In 1957, when the project obtained its first funding (a grant for use in planning) John Flanagan was living in Pittsburgh, where AIR's headquarters were at the time. He appointed Robert Craig (a member of the AIR staff at the time) as Project Director. The project was then a very small one; the planning activity was virtually a one-man operation. In the spring of 1958, when Bob Craig resigned to join the faculty of Marquette University, John Dailey, who had been Technical Director of the Research Division of the Bureau of Naval Personnel, resigned from that position to become Project Director of what was later to become "Project TALENT". At that point, since Bob Craig's departure left no project staff in Pittsburgh except the Principal Investigator, who habitually spent part of every week in Washington anyhow, and since the new Project Director was in Washington, project headquarters transferred to that city. Within a few months the project staff enlarged considerably from its one-man size, and it has never again shrunk that small (although its size has fluctuated wildly from time to time). Four members of AIR's research staff transferred from other activities to the new project. They were Dorothy S. (Jo) Edwards, William A. Gorham, David B. Orr, and the author of this History.

The geography got complicated about 1959, when Project TALENT started to need a computer. Though the project staff (by then, a fairly sizable group) was in Washington, it became clear that the computer we were going to have to use would be in Pittsburgh (at the University), and for a very good reason. Because the Cooperative Research Grant was to the University of Pittsburgh rather than to AIR, the Project would have free use of the University's computer. Free computer time is a nice thing to have, but as we shall see (in Chapter 13) the distance of 250 miles between the location where the Project Director and most of the project staff were and the University's Cathedral of Learning, where the computer was ensconced, caused problems. In 1964 these particular problems were resolved by moving project headquarters back to Pittsburgh (where they had started out). Only two staff members (Marion Shaycoft and Dick Williams) made the move to Pittsburgh. The rest all left the project (either for other activities at AIR or to transfer elsewhere). For the next two years the Principal Investigator, the Project staff, and the computer were all based in the same city--Pittsburgh.

In the summer of 1966 the Principal Investigator moved across the country to Palo Alto while the rest of the project remained in Pittsburgh for another year. New OE regulations that permitted a nonprofit organization such as AIR to receive direct support for Project TALENT without the necessity of using the University as an intermediary led to a formal severing of the Project's relations with the University, the transfer of the contract to AIR (effective the beginning of 1967), and a half-year later, another move of Project headquarters--this time to Palo Alto, where the Principal Investigator already was. Since the severing of the Project's ties to the University meant that we no longer had access to virtually unlimited computer time on the University's computers, the computer operation as well as the rest of the project could make the move to Palo Alto in the summer of 1967. The project (including Principal Investigator, staff, and computer) has been there ever since.

The geography of Project TALENT--in other words who and what were where when--is summarized in Table 9-1.

TABLE 9-1. The geography of Project TALENT
(Who and what were where when)

	Computer and Computer Staff	Principal Investigator	Other Project Staff
1957-58	---	Pittsburgh, commuting to Washington weekly	Pittsburgh
1958-59	---	Pittsburgh, commuting to Washington weekly	Washington
1959-64	Pittsburgh	Pittsburgh, commuting to Washington weekly	Washington
1964-66	Pittsburgh	Pittsburgh	Pittsburgh
1966-67	Pittsburgh	Palo Alto	Pittsburgh
1967-76	Palo Alto	Palo Alto	Palo Alto

Chapter 10. EARLY REPORTS

In this chapter we discuss, very briefly, the research reports which were produced by the TALENT staff up to 1966. Chapter 20 ("Later Studies") does the same thing for subsequent reports. Admittedly the division point between "early" and "later" is to a great extent arbitrary.

However the reports discussed in the present chapters are ones which, it is hoped, provide a basic understanding of the project, the data, their scope, and their potential. (Only major reports are discussed here, not short articles, or papers presented at professional meetings.) In neither Chapter 10 nor Chapter 20 are the findings of the studies discussed, since that is beyond the scope of this History.

A. Basic reports

Five reports fall in this category. Two of the five are introductory reports, presenting an overview of the plans for the project, and describing the preliminary phases. The first of these two is "Project TALENT: Designing the Study" (Flanagan, Dailey, Shaycoft, Gorham, Orr, and Goldberg, 1960, 448 pages). The second is a considerably abridged and re-edited version that has been published by Houghton Mifflin, as a hard-cover book, under the title "Design for a Study of American Youth" (Flanagan, Dailey, Shaycoft, Gorham, Orr, and Goldberg, 1962, 248 pages).

The other three of the "basic" reports deal respectively with school characteristics, high school student characteristics, and the results of the study of the age 15 sample. These reports are, respectively:

Studies of the American High School
(Flanagan, Dailey, Shaycoft, Orr, and Goldberg, 1962, 375 pages)

The American High School Student
(Flanagan, Davis, Dailey, Shaycoft, Orr, Goldberg, and Neyman, 1964, 738 pages)

Studies of a Complete Age Group--Age 15
(Shaycoft, Dailey, Orr, Neyman, and Sherman, 1963, 288 pages)

B. Equating studies

The Air Force and the Navy both gave us contracts to equate their tests to the TALENT battery, and perform related analyses. The resultant reports are, respectively:

For the Air Force:

Calibration of Air Force Selection Tests to Project TALENT Norms
(Dailey, Shaycoft, and Orr, 1962)

Calibration of academic aptitude composite to Project TALENT results
(Shaycoft, 1963)

For the Navy:

Comparisons of Navy Recruits with Male High School Students on the Basis of Project TALENT Data
(Shaycoft, Neyman, and Dailey, 1962)

C. Subsequent studies applying the results of the equating

Following the equating of the Air Force tests to the TALENT Battery, the Air Force commissioned us to apply the results of the equating in various ways. Some of the resultant reports are:

Comparison of Air Force Recruits with Segments of the High-School-Age Population
(Shaycoft, 1963)

Normative Distributions of AQE Indexes for High-School-Age Boys
(Tupes and Shaycoft, 1964)

Chapter 11. RETURNING TO THE SCENE

The original plans for Project TALENT had included nothing about retesting. But during the school year 1962-63, the year that our original 9th-grade cohort was due to graduate, we decided that it would be highly desirable to retest some of them; it would be our last chance while they were still in school. And we felt that much could be discovered about the factors that affect the amount of learning and other changes that occur in high school.

A. Designing the retest study

Having decided to retest, the only questions remaining were who, what, where, when, how much, and how.

1. Who

Obviously we were limited to members of the 9th-grade cohort, since they were about the only members of the original sample who were still in school. But how many? We decided to aim at testing in roughly one-tenth of the original schools in our probability sample and to test all 12th-graders in those schools.

2. Where

It was decided that not all categories of schools should be included. Nonpublic schools would be omitted, and because there were indications that it would be very difficult to get permission to retest in the five largest cities it was decided not to include those cities (except for some vocational high schools).

Those exceptions eliminated four of the 19 categories in the taxonomy of high schools developed by John Dailey and first reported on in "Studies of the American High School" (Flanagan, Dailey, Shaycoft, Orr, and Goldberg, 1962); but 15 categories remained. It was decided to select the sample of schools by stratifying into those 15 categories, to insure that all of them were represented. Accordingly, a sample of 144 schools was selected. This was a little more than we wanted, but we anticipated that some of them would decline. Some of them did, but not as many as we had expected--so we found ourselves with 108 schools to test. This was a somewhat larger group than we had anticipated, but we were pleased that the discrepancy was in that direction, not the other.

3. What

The question we had to answer was: What tests and questionnaires shall we retest with? We decided that we wanted to use almost the entire original battery. The only instruments that would be omitted were the Memory for Sentences test, the Preferences test, and the Student Activities Inventory (the personality inventory). We also decided to revise and shorten the Student Information Blank.

4. When

The best time for the retesting would be the spring, so that the data would be comparable, in regard to time of year collected, with the 1960 data.

5. How much

In contacting the schools selected for the sample, we had not asked for two days of testing time (as was done for the 1960 testing) but for just one day, since we felt schools would be readier to participate on that basis. This meant that each student could take only about half the tests in the battery.

6. How

We didn't want to just divide the battery in half and give half the students one half and the other half the other half. That would have meant we couldn't get correlations between the tests in one half and those in the other half. Because we felt it was necessary that every test be taken in conjunction with every other, we divided the battery in four parts (the four original half-days) and paired every one of the four resulting half-day half-batteries with every other, to get six full-day batteries (Retest Batteries A, B, C, D, E, and F). For good measure (no pun intended), we put the Abstract Reasoning Test in all six batteries, to provide a common metric. Each of the batteries was administered in about one-sixth of the schools (except that a seventh battery--Battery V--was used in 10 vocational high schools, added later).

B. Processing the answer sheets

Things had progressed rapidly at Measurement Research Center since their processing of the answer sheets from the 1960 testing. By 1963 they had a document-to-tape optical scanner that could handle the kind of answer sheets that in 1960 had required the Scoring Machine; the new optical scanner would put every item response directly on tape. From the tapes containing item responses, we could thus obtain whatever scores we wanted, using our own (or rather, Pitt's) computer to calculate them.

This meant that in addition to studying change over the 3-year period, we would also be able to obtain split-half reliability coefficients where appropriate and we could also, at last, get empirical data on the degree to which the tests were speeded for 12th-grade students.

C. Data analyses and final report

The data were analyzed in many different ways, and from many different angles. Several methodological innovations were introduced for analyzing the data. The final report, "The High School

Years: Growth in Cognitive Skills" (Shaycoft, 1967, 376 pages), discusses both the methodology and the substantive results in detail.

D. Test reliability data

One important point about the report mentioned above is that at present it is the best source of test reliability data. Columns 5 and 6 in Table 4-7 (Shaycoft, 1967) contain split-half reliability coefficients for the grade 12 norms group. Researchers using TALENT variables and needing reliability coefficients are urged to use the values found in that table; but it is suggested that they apply the conventional correction-for-range formula to obtain reliability estimates applicable to the set of cases on which the corresponding research data (test scores) are based.

It might be mentioned, incidentally, that though some of the tests are very short (since they were designed for research use, not for decisions on individual students) and may therefore have only moderate reliabilities, the reliabilities are high for most of the tests, and on a per-item basis they are, on the whole, excellent, even for the shortest tests.

Chapter 12. THE 4% SAMPLE

When we were arranging with MRC for the processing of the answer sheets to be used in the 1963 retesting discussed in the previous chapter, we found out about the advances in MRC's equipment (see Chapter 11, Section B). These advances meant that the same answer sheets used in 1960 and processed on the Scoring Machine, in other words answer sheets A and C, could now be processed on an optical scanner which would produce output containing individual item responses, rather than merely 13 scores. Furthermore the output would be tape, not cards; and thus only one pass of an answer sheet would be necessary to get all the item responses recorded.

We decided to take advantage of this improved equipment by excavating answer sheets A and C for our original 4% sample--i.e., cases for which the six-digit student identification number ended in 00, 25, 50, or 75--from the catacombs where they were stored, and shipping them to MRC, where we could have them processed very inexpensively on the new equipment, by arranging for them to be "piggy-backed" on the corresponding sets of answer sheets from the 1963 retest. (It was not necessary to have any 4%-sample answer sheets other than A and C reprocessed, since, it will be recalled, the other answer sheets from the 1960 testing (answer sheets B1 and B2, and Record Form Z) had been in a different format and had originally been processed on the Document Reader then existing, which had produced card output containing individual item responses. (The Document Reader had been unable to handle the format used for answer sheets A and C, which had been designed for the Scoring Machine.)

Thus for 4 percent of the cases we would be able to develop a tape containing every item response, as well as the standard set of scores, augmented by whatever additional scores we might wish to obtain either for experimental purposes (including methodological research) or because we considered them useful for substantive research.

To produce this magnificant data tape containing the complete set of data would require that the new tape containing the answer sheet A and C data be merged with the item responses and other data from answer sheets B1, B2, and Z already in the tape master file. This would not be an easy job. At that point our tape files were in a rather primitive stage of development--not even in perfect numerical order--so that merging could not be handled efficiently. Furthermore not everybody's record had as yet been incorporated in the tape master file. The same difficulty in merging would apply of course to the retest data tapes.

It was decided, therefore, that the sensible thing to do would be to lay aside the new tapes received from MRC (i.e., the 4%-sample tapes for answer sheets A and C and all the 1963 retest data) until the master file tapes were in better shape, before proceeding with efforts to merge and edit. Unfortunately, by the time the master file was getting into fairly usable shape the project had moved to Pittsburgh and no longer had the same Project Director it had had in Washington. None of the subsequent Project Directors agreed that a complete and fully edited 4% sample data tape would be a valuable asset to the project--and so they never assigned this task high priority; they never assigned it even medium priority; it was always shunted to the bottom of the list. And there it sat--for years! Occasionally some member of the programming staff would be assigned responsibility, at a low level of priority, for some sort of superficial editing and formatting, but nothing much ever came of it.

The author of this History continued to campaign (incessantly, some might say) for getting the 4%-sample tape file in first-rate shape. Thus when she had an opportunity, about December 1973, to write a proposal to be submitted to NIE (at their invitation) to revamp the TALENT tapes and their documentation (see Chapter 13, section _) she made certain that fixing up the 4%-sample tape and documentation was an integral part of the work proposed. It was funded in the summer of 1975. For further details see Chapter 13.

fill in

Before dropping the subject of the 4%-tape even temporarily we should point out that not the least of its advantages would be that it would yield eminently sound reliability estimates for any variables for which internal consistency estimates (e.g., split-half reliability, KR-20 reliability) would be appropriate. For many of the TALENT variables (including most of the test score variables) there was no way of determining reliability on the basis of the regular 1960 data. The 1963 retest data would to a certain extent share these advantages with the 4%-sample data, in regard to the determination of reliability, but the 4% data would be preferable for the following reasons:

1. It would be available for all four grades, not just for grade 12.
2. It would be for a probability sample of the grade.
3. It would be available for all variables in the TALENT battery (including some that hadn't yet been computed, and even some that hadn't yet been defined) rather than being limited by the composition of the six retest batteries.

Chapter 13. TAPES: TRAUMA, TROUBLES, TRAVAIL, TRIALS, TRIBULATIONS, AND T.E.A.R.S.

A. TALENT's tapes: Ancient history

1. Iowa City (1960)

When the original answer sheets had been processed at MRC in 1960 the tapes didn't yet exist, but about 5 million IBM cards (11 per student) did. One of the limitations of the optical scanning equipment that was used was that unless we had arranged to get counts of "number of items attempted" on every test* rather than merely counts of the number answered correctly, there was no way that it would be possible, on the basis of the output cards alone, to distinguish between zero scores, denoting that the test was taken but no item was answered correctly, and missing data (no item attempted, suggesting, though not indicating with certainty, that the student had been absent from the testing session). Both missing data and genuine zero scores were represented on the IBM cards by zero punches.

2. Coping with geography

It will be recalled (see Chapter 7 section E3, and Chapter 9) that when the answer sheets were being processed at MRC we (most of the project staff) were in Washington, while our computer staff and the computer we were using were both in Pittsburgh--about 250 miles away. This sort of splitting of the operation is really a bad idea, but it was a logistic necessity. We had to use the University of Pittsburgh computer because we were getting virtually unlimited free time on it. However I would like to offer this advice to future researchers: If you have any choice at all in the matter, the computer staff should be at the same site as the other staff members--and the computer should also be nearby--preferably within a few miles and certainly no more than 15 or 20 miles away. (Not everything in a project of the magnitude of Project TALENT can be handled reasonably on a terminal.)

---

*For budgetary reasons doing this would have been out of the question, since it would have doubled our scoring costs.

3. Zeros and blanks

The fact that the output of the MRC Scoring Machine would not distinguish between zeros and blanks had been foreseen. We (in Washington) were aware of the Scoring Machine's limitation in this respect, and had intended it to be taken care of in the tape editing process, by making the assumption, in the case of answer sheets that had been processed on MRC's Scoring Machine, rather than the Document Reader, that if a student's scores were zero for all the tests administered in a half-day testing session, the student had been absent that half-day, and his "zeros" should be converted to whatever code was being used to indicate "blank" or "missing data". But presumably because of the 250 miles separating that intention (in the heads of some people in Washington) from the programming operation (by personnel in Pittsburgh) the distinction between zeros and blanks didn't get edited into the tape. In the years that followed, sporadic efforts were made to get the necessary corrections made, and this was largely accomplished in the case of students who lacked an entire answer sheet which might (or might not) indicate absence for a full day, but not for students who missed only a half day and were not missing an answer sheet.

While this problem applies to a very small number of cases (certainly less than 1/2%) it has been a continuing concern. Recent efforts at reducing or eliminating such problems are discussed later in this chapter.

4. Building a tape master file

Zeros and blanks (and the difference between them) didn't constitute the only problem. Far from it! From some of the answer sheets two cards were supposed to have been made, and from others, three. But when the time came to merge the tape records for cards which had been made from the same answer sheet, and which therefore should have matched perfectly, there were lots of unmatched records. And when, because of double-punched columns, two tape records were made from the same card (see Chapter 7 sections E2 and E3) there should have been perfect matching between the two. Inexplicably, there wasn't. (The word "inexplicably" is not intended to imply that there was no possible explanation. As a matter of fact several explanations were possible; but because of inadequate control procedures, there was no way of knowing which was the correct one.) Of the 17 sets of tape records no 2 matched perfectly. And again, partly because of the lack of good communication between the Washington staff and the programming staff in Pittsburgh, bad procedures were being used and bad results were being obtained, in the efforts to merge the 17 tape records for each student to create a "master file" containing the students' complete records. The master file had to be created first for "school subsample 0" as indicated in Chapter 3 section A4, so that we would have a 10% sample on which to do preliminary analyses. The remaining 10 school subsamples (9 probability subsamples and one non-probability) could of course have been processed all together--each step being done for all ten at once. However, this would have greatly increased the costs of coping with hardware errors, a not infrequent problem in those days. Instead, it was decided to continue on the one-subsample-at-a-time routine--although this meant that each of the very numerous steps was carried out 10 times instead of just once. And then it turned out that the 4% sample, which had been given special treatment initially so that preliminary norms could be obtained for use in reports to the schools, had been omitted from subsequent processing; this meant that more backtracking was necessary to prepare master file records for these cases. Small wonder that the master file production seemed endless! It was endless (figuratively speaking). It went on for years (literally).

Some of the cases that should have merged didn't, because of either bad card records or bad tape records produced from good card records, or bad sorting. (The latter could wreck the merging not only of a missorted case but of every case subsequent to it in the same file—a fact which was ignored for far too long, resulting in far too many failures to merge and far too many unnecessary steps.)

By sometime in 1963 all the merging operations had been completed at least once, resulting in merged master file records for 292,050 cases in the grade 9-12 probability sample, plus some other cases, together with unmerged remnants for 100,000 cases or more. The merging efforts had involved, primarily, matching on student identification number and beginning of last name.

Efforts to merge what data there were for 100,000 cases represented in the remnants was a more complicated operation, since many of the failures to match were due to errors that had been made in position-coding a digit of the student identification code or a letter of the name, on the answer sheet, or to failure of the optical scanner to pick up one or two characters of the identification data. Thus visual inspection of printouts of the unmatched cases, inferences as to what was wrong, hand-matching, and keypunching the correction cards were all integral parts of the ever-so-laborious, ever-so-time-consuming process of converting cases in the remnant filed to cases in the master file. (In retrospect, a major part of this could have been obviated through different procedures in the earlier phases--but that is water over the dam.)

Efforts to merge the 100,000 or more cases represented in one or more of the various remnant files were continued, and were still going on (with steady progress being made) after the project moved from Washington to Pittsburgh in 1964; and it was still going on after the project moved from Pittsburgh to Palo Alto, in 1967! But before that happened, a part of the problem was eliminated by the Project Director, who ordered that certain tapes which contained unmerged or partly merged "remnants" (in some cases as many as about 20,000 cases*) be scratched. This was an unfortunate "solution," as many of these unmerged or partly merged remnants contained solid sets of data in their own right--for instance all of the Information Test scores, of which there are about 40. But getting rid of these remnants certainly made the tape situation seem neater. The order to scratch remnants did not apply to all of them. Some were retained, for future efforts at merging.

After the move to Palo Alto, Dick Williams, who had become the Director of Computer Operations, worked doggedly on the "remnants of the remnants"and made excellent progress. He was able to clean up most of the remaining remnants, to the point where what was left unmatched could be regarded as negligible. Dick deserves the credit for getting 83,148 probability sample cases (plus some other cases) added to the master file, thus bringing the probability sample total up to 375,198 and the grand total for grades 9-12 up to 404,218. Even by today's standards this is a very considerable accomplishment!

The point of this discussion is that even though data processing technology has improved vastly since the original answer sheets were scanned, researchers are frequently unversed in proper data control procedures. When a project is the size of TALENT, poor procedures can cause very large headaches that are difficult if not impossible to correct.

---

*If memory serves

5. Design of the tape master file

Some readers may wonder why this section doesn't precede the previous one, entitled "Building a tape master file". Preferably, designing a tape should precede building it, not follow it, but it doesn't always happen that way. The data in the master file were not arranged on a logical basis that would make the format easy to understand and the tape easy to use--the format was not logical. Rather, it had just "happened". And that has been another continuing source of problems over the years. Around 1965 one of the programmers was apparently given the assignment of revising the format and, of course, the tapes--but the changes made were rather superficial and hardly improved matters at all. However, that was the format that existed until recently. As we shall see, a major effort to overhaul the tapes and improve their documentation was initiated in 1975 and completed by mid-1977.

6. Other instances of data loss

There were other instances over the years in which decisions were made that resulted in some data being lost or destroyed. While the loss of any data is unfortunate, the need to have a manageable data system and the presence of cost and other constraints on what could be done must be recognized. These instances are briefly described below to illustrate the kinds of problems encountered.

Data about non-examinees. It will be recalled from Chapter 5 that teachers were requested to fill out Record Form Z for absentees and other students who weren't tested, and to position-code the reason for their not having been tested. Regional Coordinators were requested to do the same thing for 15-year-olds not in grade 9, 10, 11, or 12. Unfortunately, these data no longer exist.

Redundancy reduction. A lot of redundancy had been built into the set of data collected from the students in 1960. This was done deliberately, as insurance against undetected error in our records of important bits of data (e.g., sex, grade, and age, all of which were recorded in at least two different ways, sometimes on different answer sheets). The idea was that if the two position-codes for, say, sex, agreed we could have confidence in the response. If they disagreed one of them would obviously be wrong but we wouldn't know which. The problem might be resolved, however, by inspection of a printout showing first names, although this is a very expensive procedure. An alternative would have been to have a code to indicate that both answers had been given, and then to wait for the follow-up to resolve the dilemma. What happened, instead, was that an arbitrary decision was made as to which of the two sources of the student's sex would be used (e.g., Record Form Z), and the other source (e.g., Answer Sheet A) was ignored. The same thing happened with the record of what grade the student was in.

Conversion to the IBM 360. In 1967, when we were getting ready to move to Palo Alto, the data processing staff was busy converting data tapes so that they would be usable on the 360 we would be using in Palo Alto, rather than the 7070 at Pitt. Since the workload of the data processing staff was heavy, the Project Director decided to lighten it by having them "scratch" (wipe out) certain data tapes rather than convert them. The principal tape file chosen for this form of oblivion was the one containing the data for 15-year-olds not in school. The justification that was offered for this action was that the report on 15-year-olds had already been written, so nobody would have any use for the data. It is fortunate that this irreplaceable data was a small file, not a large one.

In addition the file of 15-year-olds in school was scratched, on the grounds that it could be reconstructed from the 15-year-olds' duplicate records on the appropriate tapes organized by grade and sex. On similar grounds the Project Director ordered that on the tape containing the 1963 retest data and the 1960 data for the same cases, only the 1963 data be converted. His explanation was that in Palo Alto the retest data could again be matched with corresponding 1960 data. As it turned out, after the editing and correcting of the "age" variable in connection with the tape editing and reformatting (see Section D below), the 15-year-old file needed to be reconstituted anyway.

Nonrespondent data. In the early follow-up surveys, information on the special sample nonrespondents (e.g., the reason for nonresponse) was not added to the data tapes and was subsequently scratched. Beginning with the 5-year follow-up of the 11th grade cohort, information on special sample nonrespondents has been included in the final data tapes. The information has proved essential to the new procedure for adjusting weights to correct for nonresponse bias (see Chapter 15).

Mailer tape

Some follow-up data has gotten lost because it was obtained in connection with efforts to update the mailer file. In this category is some of the information that would cause deletion of a sample member from the mailer tape. When sample members were flagged for deletion the reason (e.g., reported death--surely a relevant item of follow-up information) was not transferred to the data tapes. Also it is not clear to what extent records have been retained, on tape, of old names and addresses after their replacement on the mailer tape by newer ones. The outdated information needs to be preserved for use in follow-back studies (see Section C below); but not all of it has been.

B. Interim Alphabetic Directory

Despite the title of this chapter we certainly do not want to give the impression that everything that was done in collection with data processing was done badly. Many things were done well. The alphabetic directory was one of them, even though it does not meet current needs well. The alphabetic directory was prepared about 1963. It is a printout containing names of members of the TALENT sample in alphabetic order, together with other identigying information, and one of its intended functions is for use in follow-back studies, which are discussed in paragraph C below.

When the alphabetic directory was prepared in 1963 or thereabouts, the programmer quite properly designated it an "interim directory." He was aware that the master data files from which he prepared it were at that stage far from complete. It will be recalled that in late 1963 the master file, which now containes our 375,000 cases, contained only about 292,000 cases. The alphabetic directory created then may have contained even fewer probability-sample entries,* certainly it contained no more. But it served its immediate purpose--which was to function as a stopgap.

Unfortunately that "interim" directory is still our only directory, 13 years and probably 90,000 cases later. It is an enormous thing--about 34 side-inches of printout bound in 23 volumes, with inelegant but sturdy and effective covers made out of printout-sized (i.e., 11" x 15") rectangles of the kind of brown cardboard used in the manufacture of ordinary cartons; each volume is held together at the spine by a "home-made" binding of gauze and glue.

Although the alphabetic directory was fine when it was created it now suffers from the following deficiencies:

a. It lacks some unknown number of entries, but probably at least 90,000; possibly more.

b. For the purposes for which it is used, it would be more effective if each entry contained more information about the individual.

c. After 13 years of very hard use its physical condition is deteriorating badly. Parts of it have almost disintegrated.

---

* It contains "nonprobability cases," as well.

A complete update was planned several years ago to include not only a new alphabetic printout containing more information and formatted for good readability, but also a directory tape containing the same information, and also a printout in numerical order, which we also sometimes need. Unfortunately, overall project cost constraints have precluded updating the directory and preparing a directory tape.

C. The concept of follow-back studies

The follow-back technique is designed for studies in which the interest is in members of the TALENT sample who subsequently became members of some specific group or category--for instance people who went to prison; or graduated from medical school; or became patients in a mental hospital; or played professional football; or won some kind of award; or were elected to state legislatures; or were members of any other delimitable group that might be of interest. The follow-back procedure is simple. The researcher provides a list of the members of the group he is interested in, who happen to be in the right age range to have been in the TALENT sample. He also provides auxiliary identifying information if possible--for instance date of birth, high school attended, maiden name if female, father's first name, etc. If the age range is right, presumably about 4 percent of the group might be in the TALENT sample. A search is therefore made of the TALENT tapes to locate the tape records of anyone on the list provided who also was in TALENT. (The tape search is usually supplemented by a search of the alphabetic directory of TALENT sample members). Data from the TALENT records of the cases identified as members of the group that is of interest in the particular study are then analyzed.

This follow-back technique is a potentially powerful one. But primarily because of the inadequacies of the alphabetic directory it has not worked as well as in theory it should. In some studies only about 1½ percent of the cases in the right age range have been found, instead of the expected 4 percent.

D. Tape Editing and Reformatting System (T.E.A.R.S)

The Tape Editing And Reformatting System was a major effort designed to resolve many of the problems with the data files referred to above. The production and documentation of the reformatted files was a significant accomplishment and greatly enhanced the usability of the data. The activity started in 1975 but the idea for it, or to be a bit more precise, the idea for a major part of it, goes back at least 10 years to 1964 or 1965, when the author of this History, having moved to Pittsburgh and thus having gotten a little closer to the computer operation, perceived an urgent need for reorganizing the tapes (see Section A5 above) and improving the documentation, and began campaigning for it. Subsequent developments are described below in four phases: (1) the NIE Panel; (2) the proposal; (3) September 1975 - April 1976; (4) April 1976 - May 1977.

1. The NIE panel

Since most of the project directors were more concerned with research issues than with the need for improving the tapes and their documentation, nothing happened until late in 1973, when NIE brought in a panel of outside consultants, one of whose functions was to review Project TALENT's proposal for continued funding (see Chapter 17, Section C).

Most of the panel members had some prior familiarity with Project TALENT, and at least one had been a Data Bank client (see Chapter 22). The panel made a recommendation along the exact lines that Marion Shaycoft had been urging; they recommended a thoroughgoing reorganization and documentation of the data tapes. The recommendation seemed to imply, also, although it did not so state explicitly, an updated Data Bank handbook was needed.

2. The proposal

As a result of the Panel meeting, the project monitor requested the author of this History to draft a proposal for tape improvement and documentation. She immediately (and enthusiastically) complied with this request. The proposal was apparently tabled for about a year, but in the spring of 1975 we were asked to submit a formal proposal. The proposal Marion Shaycoft prepared in response to this request organized the work in separate modules with built-in priorities, as shown in Figure 13-1.

A request was received form NIE to cut the scope of the proposed work for the first year, because of budgetary limitations. The modular arrangement made this easy to do. In the revised proposal the following modules were retained:

#1a.	1960 master file	Reorganization, expansion, and documentation of these tapes
#1d.	4% sample	
#1b.	Follow-up tapes	
#1c	Retest tape	
#5	Data Bank Handbook	

Figure 13-1 contains additional details.

The staff had to be cut a bit below the Panel's recommendation of one full-time director and two full-time programmers.

For administrative reasons the proposal was incorporated in a larger proposal that also covered other aspects of Project TALENT. It was funded, with Marion Shaycoft as Director of this part (i.e., the tape improvement and documentation , and the preparation of the new Data Bank handbook), and with Laurie Wise and Wendy Bartlett as her colleagues on this activity. It was at this time that Laurie Wise developed the acronym for the task, T.E.A.R.S.

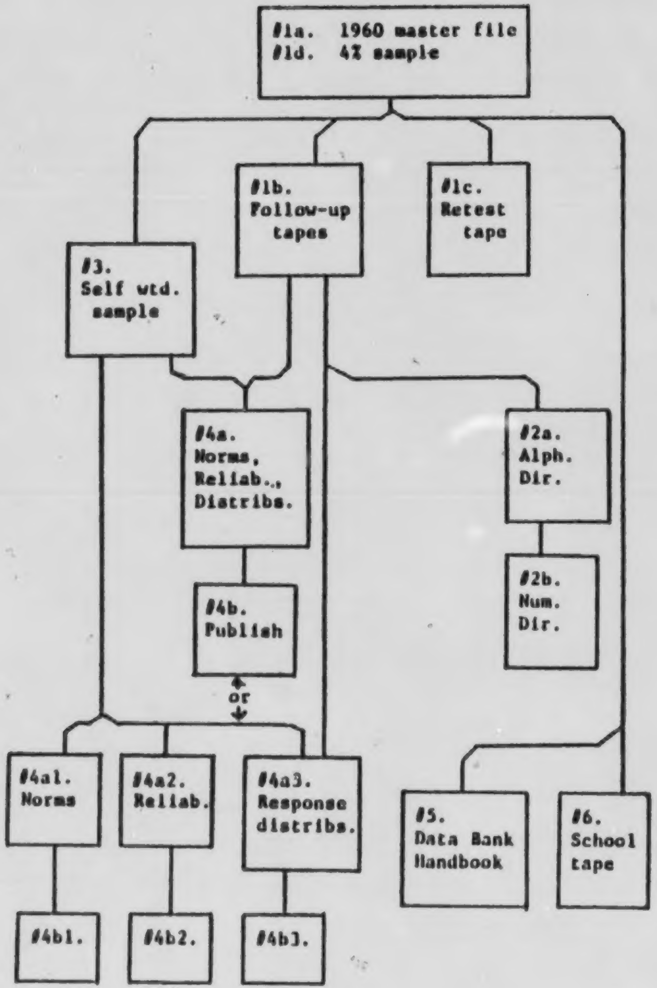
3. September 1975 - April 1976

Work began on the project September 1975.* Marion Shaycoft drew up plans for formatting which would, in her opinion, meet the necessary standards for good documentation. The principle she followed was that in documenting computer-related information for the use of researchers who are not necessarily programmers, a necessary (but not sufficient) condition for good documentation is that the answer be affirmative to all three of the following questions:

---

* The Data Bank handbook aspect of the project is discussed in Chapter 22, Section F.

Figure 13-1. Priority chart* for each additional step**


#1a. 1960 master file
#1d. 4% sample
#1b. Follow-up tapes
#1c. Retest tape
#3. Self wtd. sample
#4a. Norms, Reliab., Distribs.
#2a. Alph. Dir.
#4b. Publish
#2b. Num. Dir.
or
#4a1. Norms
#4a2. Reliab.
#4a3. Response distribs.
#5. Data Bank Handbook
#6. School tape
#4b1.
#4b2.
#4b3.

*No step should be undertaken unless every step to which it is connected by a line and which is above it in the chart is also undertaken.

**The step numbers shown in the above chart have the following meanings:

1. Modification and documentation of student data tapes
    a. Master file for 1960
    b. Follow-up tapes
        1) 1-year follow-up tapes (containing 1-year follow-up data and 1960 data)
        2) 5-year follow-up tapes (containing 5-year follow-up data, 1-year follow-up data if available, and 1960 data)
        3) 11-year follow-up tapes (containing 11-year follow-up data, 5-year and 1-year follow-up data if available, and 1960 data)
    c. Retest tape

        This tape should contain both the 1960 master file data for the students when they were in grade 9 and the grade 12 data obtained in 1963.
    d. 4%-sample tape
2. Directories and directory tapes
    a. Alphabetic
    b. Other
3. Self-weighted samples for 1960 master file
4. Norms, item response distributions, and reliability coefficients
    a. Compute those that haven't been computed
        1) Percentile norms, with means and standard deviations, for all test variables, a priori composites, and inventory scales
        2) Reliability coefficients
        3) Item response distributions for inventory and questionnaire items
    b. Prepare for publication
5. Revision, expansion, and updating of Data Bank Handbook
6. Modification and documentation of the school data tapes

a. Is the documentation correct?

b. Can it be read easily and understood fully by someone who is adept at reading English but illiterate in computerese?

c. Does it give him or her all the information he or she needs?

The answers to the following questions also have to be affirmative if the documentation is to serve the needs of the programmer:

a. Can the documentation be used easily by a competent programmer who has little or no knowledge of research?

b. Can he use the documentation without having to ask a lot of questions of the person who prepared it?

c. Can he proceed with his work without having to ask the person or persons who supplied the tape that is documented questions about it?

To meet these standards the plan called for documentation in three parts.

Part A was to be as described in Figure 13-2.

Part B was to give the information about what kind of information is on the tape, how it is represented on the tape, and where on the tape it is. More specifically, it was intended to provide full details about what information is in each byte position; exactly what the codes are (both on the source documents and on the tapes; how omitted items (in case of item response data) are handled; what representation is used for missing data; how many decimal places there are; what the possible range of the scale is for any variable; and any references and other explanatory notes that seem necessary or useful.

Several forms, all of them variants of the same general form, were drawn up for use in Part B. The idea was that the particular form used would depend on the general nature of the material to be formatted on that sheet—e.g., a different form would be desirable for questionnaire item responses from the one designed for indicating scores on test variables.

Figure 13-2. Description of Part A of tape documentation.

Part A will describe the general characteristics of the tape, including, among other things, such features as:

1. Physical characteristics, e.g.,
    a) seven-track or nine-track
    b) number of bytes per inch (bpi)
2. Physical record length
3. Block size
4. Kind of notation used (e.g., EBCDIC, packed binary, etc.)
5. Cases (or logical records) included in the file — described fully both in terms of inclusion and (insofar as necessary for clarity) exclusion.

    For instance the cases included in one of the 13 to 16 tape files to be prepared under the present grant that consist of 1960 master file records might be described as: "All Grade 11 males in the Grade 9-12 probability sample." Another of the 13 to 16 tape files — the one briefly described as the "nonprobability" file — would be more fully described as "all cases included in the 1960 TALENT testing who are not included in any of the following files:

    a) the grade 9-12 probability sample file
    b) the age 15 probability sample file
    c) the Knoxville-and-Knox County saturation sample file
    d) the grade 8 semi-probability sample file."

    [The term "grade 8 semi-probability sample" will be fully explained in the documentation for that file.]

6. Order in which the cases (or logical records) are arranged. For instance, the arrangement of cases on a tape file might be described as "by student identification number, low to high."

    Another arrangement might be described as follows:

Level	Characteristic on which arrangement is based	Order
MAJOR CLASSIFICATION	Grade	Low to high
INTERMEDIATE CLASSIFICATION i.e., classification within major classification	Sex	1, 2, Blank
MINOR CLASSIFICATION	Student Name	Alphabetic

Figure 13-3a shows a Part B form intended primarily for test scores; figure 13-3b shows a Part B form for other kinds of information including questionnaire items.

Part C of the tape documentation was to consist of whatever appendices are needed to round out and complete the documentation. Appropriate parts of the appendices were to be referenced at appropriate points in Parts A and B. It was anticipated that in both Part A and Part B it would be necessary to add a considerable amount of cross-reference information (to appendices, etc.) for more information about variables, coding systems, specific codes, etc.

The principle was established that instead of spending 8 or 9 minutes planning the tape preparation and 8 or 9 years executing it, as seemed to have happened in the original preparation of the master file tapes,* we would spend at least three-quarters of the available time planning and the remaining quarter executing the plan. This meant among other things, that all the formats would be designed and documentation completed before the new tapes were made.

In December 1975 a 100-page packet of materials, consisting of fairly detailed plans for the tape revision and documentation and a sample format, were sent to three former users of the Data Bank services, whom we had invited to serve as consultants. The three consultants were:

John Hause (University of Minnesota)
Lloyd Humphreys** (University of Illinois)
Christopher Jencks (Harvard)

The reviewers reactions were very favorable.

Things proceeded along those lines until April 1976, when the

---

* See paragraph A5 above and consider the fact that cases were still being added to the master file about 1969.

** Lloyd Humphreys was unable to review the materials when we sent them, but each of the other two consultants arranged to have someone knowledgeable in addition to himself look at them so that in effect there were four active reviewers.

Project # ______ Project: ______________________________

TAPE FORMAT for ______________________________ LOGICAL RECORD Page B-____ of ____ Date ____________
Mo./Day/Yr.

ition	Field #	Characteristics of Field				DESCRIPTION			Contents of Field			References	Miscellaneous Notes
		Length	Alph or Num. Mode	Kind of Data	Type of Scale	Variable #		Missing Data Code	No. of Dec. Pl.	Possible Range			
										LOW	HIGH		

Figure 13-3a

C - 2a-h 111975

TAPE FORMA ______________________ **LOGICAL RECORD** Page B-__________ of ______ Date __________
No./Day/Yr.

		Characteristics of Field				DESCRIPTION			CODES		Contents of Field				
												Possible Range			
Position	Field #	Length	Alph or Num. Mode	Kind of Data	Type of Scale	Scale	Item #		Code for Omitted Item	Missing Data Code	No. of Dec Pl.	LOW	HIGH	References	Miscellaneous Notes

Figure 13-3b

Director of Project TALENT made some administrative shifts in personnel. To give Marion Shaycoft more time for the other major responsibilities she then had*, and still permit completion of T.E.A.R.S. (the tape improvement and documentation activity and a new Data Bank Handbook) by 30 September 1976, Laurie Wise was made full director for the T.E.A.R.S. effort.

4. April 1976 - May 1977

Edited and reformatted tapes for the 12th grade probability cases containing 1960 data and data from all 3 follow-ups were completed in October 1976 along with a draft of the documentation for these tapes. The tapes and their documentation were put to immediate use in a study directed by JJCard of the consequences of adolescent childbearing.

Work on the data for remaining cohorts plodded along. As new problems were discovered, investigated and resolved with the testing of each new edit program, it became clear that the required programming time had been significantly underestimated. Consequently, it was decided to focus on completing the tapes of greatest current interest--the 1960 and follow-up survey data on the entire probability sample.

During the course of the project Laurie made several changes in the scope of the work reflecting changes in priority based in part on the results of initial tests. Specific changes in project activities were as follows:

ADDED TASKS:

Test record forms. A record form was filled out for each of the roughly 13,000 groups tested in 1960 containing information on the test timings, the dates tested, and any group or individual irregularities. Nothing much has ever been done with these forms. Marion Shaycoft had planned to use them to identify cases where testing irregularities might have invalidated some scores. Under her direction a preliminary plan for this was developed. Laurie Wise implemented the plans, and in some ways expanded on them. In order to correct for the irregularities noted and to add test date and group information to the individual records, data from these forms were put onto keytape, edited and corrected, and merged with the individual student data. The information from the test record forms provides an important part of the efforts to distinguish zero test scores from missing data.

---

* Writing of this History and completion of a monograph on criterion-referenced tests (Shaycoft, 1976).

Consistent follow-up formats. The follow-up data were stored on tape in questionnaire item order which varied widely from one grade to the next. The codes used for each variable also varied from one grade to the next in many cases. This meant that when the same analyses were run on different grade cohorts, vast changes in the computer programs were required because of the different formats. In addition, just finding a variable on the tape formats was difficult if you didn't already know the questionnaire item number.

It was decided that the follow-up data should be rearranged into a common format for all 4 grade cohorts. The common format developed was based on a topical ordering of the follow-up items (within follow-up) so that an item taxonomy was generated as an additional bonus.

Putting the formats online. It was decided that, insofar as possible, the new tape documentation should be stored on magnetic tape in a form that could be interpreted by computer programs as well as by people. The formats were streamlined somewhat for this purpose, as shown in Figure 13-4. Laurie indicates that the following advantages have resulted from this decision. The edit/reformat programs used the online formats in defining the input and output structures, saving considerable programmer time and reducing the possibilities for errors. Laurie also says that "The final tape format could be created from the work tape formats with a few simple text editor commands." Work tape formats for Data Bank projects have similarly been created using the text editor, saving the typing or extensive cutting and pasting that was necessary heretofore. The online text was used in generating the tables for Volume 2 of The American Citizen, saving enormous time and headaches.

Figure 13-4

```
-------------------------------------------------------------------------------------------------------------------------------
          | L |   QUESTIONNAIRE      |VARIABLE    TEXT OF QUESTION                       |    |                  |    |    |
          | E |ITEM NUMBERS AND      |  ID                                               |NO. |POSSIBLE RANGE    |    |MISS|FOR-
   BYTE   | N |   RESPONSE OPTIONS   |---------------------------------                  |OF  |------------------|OMIT|ING |MAT
          | G |   FOR EACH GRADE     | CODES  |                                          |DEC |        |         |    |DATA|LINE
POSITIONS | T |----------------------|  ON    |      TEXT OF OPTION                      |PL. |  LOW   |  HIGH   |CODE|CODE|NO.
          | H | 12 : 11 : 10 :  9    | TAPE   |                                          |    | VALUE  |  VALUE  |    |    |
          |   |    :    :    :       |        |           Q.  III  OCCUPATION            |    |        |         |    |    |
-------------------------------------------------------------------------------------------------------------------------------
          |   |    :    :    :       |04      |OCCUPATION:  CAREER PLANS                 |    |        |         |    |    |2771
          |   |    :    :    :       |        |                                          |    |        |         |    |    |
2015-2016 |  2|Q6GG:Q6GG:Q5GG:b      |0401    |Do you plan to remain on this job for the next |    |11      |       52|   0|   b|2772
          |   |    :    :    :       |        |several years?                            |    |        |         |    |    |2773
          |   |    :    :    :       |        |   5x  Yes                                |    |        |         |    |    |2774
          |   |   B:   B:   E:       |      52|  Yes, I would like to.                   |    |        |         |    |    |2775
          |   |   A:   A:   A:       |      51|  Yes, but only because I think it would be hard |    |        |         |    |    |2776
          |   |    :    :    :       |        |    for me to find another job.           |    |        |         |    |    |2777
          |   |    :    :    :       |        |   4x  Probably                           |    |        |         |    |    |2778
          |   |   C:   C:   C:       |      41|  Probably.                               |    |        |         |    |    |2779
          |   |    :    :    :       |        |   3x  Can't decide                       |    |        |         |    |    |2780
          |   |   D:   D:   D:       |      31|  I can't decide.                         |    |        |         |    |    |2781
          |   |    :    :    :       |        |   2x  No, but would like to stay in the field |    |        |         |    |    |2782
          |   |   E:   E:   E:       |      22|  Nc, I would like to change jobs, but continue to |    |        |         |    |    |2783
          |   |    :    :    :       |        |  do the same kind of work.               |    |        |         |    |    |2784
          |   |   F:   F:   F:       |      21|  No, I would like to get a job in the same field, |    |        |         |    |    |2785
          |   |    :    :    :       |        |  but at a more advanced level.           |    |        |         |    |    |2786
          |   |    :    :    :       |        | 1x  No, some other reason                |    |        |         |    |    |2787
          |   |   G:   G:   G:       |      15|  No, I would like to get a job in a different |    |        |         |    |    |2788
          |   |    :    :    :       |        |  field.                                  |    |        |         |    |    |2789
          |   |   H:   H:   H:       |      14|  No, I plan to stop work alltogether to go back |    |        |         |    |    |2790
          |   |    :    :    :       |        |  to school.                              |    |        |         |    |    |2791
          |   |   I:   I:   I:       |      13|  No, I plan to stop work altogether, for some |    |        |         |    |    |2792
          |   |    :    :    :       |        |  other reason.                           |    |        |         |    |    |2793
          |   |   J:   J:   J:       |      12|  No, for some other reason.              |    |        |         |    |    |2794
          |   |   K:   K:   K:       |      11|  No, I have already left that job since September |    |        |         |    |    |2795
          |   |    :    :    :       |        |  30, 1971.                               |    |        |         |    |    |2796
          |   |    :    :    :       |        |                                          |    |        |         |    |    |
2017-2017 |  1|Q6H :Q6H :b   :b      |0402    |Do you expect to advance in this job?     |    |1       |        5|   0|   b|2797
          |   |   A:   A:    :       |        |  Yes, I expect to advance far in this job. |    |        |         |    |    |2798
          |   |   B:   D:    :       |        |  Yes, I expect to advance some.          |    |        |         |    |    |2799
          |   |   D:   E:    :       |        |  No, I am content with my present level. |    |        |         |    |    |2800
          |   |   E:   B:    :       |        |  No, I do not expect to stay long enough |    |        |         |    |    |2801
          |   |    :    :    :       |        |  to advance                              |    |        |         |    |    |2802
          |   |   C:   C:    :       |        |  No, I have little or no opportunity for |    |        |         |    |    |2803
          |   |    :    :    :       |        |  advancement.                            |    |        |         |    |    |2804
          |   |    :    :    :       |        |                                          |    |        |         |    |    |
2018-2019 |  2|Q6I :Q6I :b   :b      |0323    |Does your job provide good financial security? |    |11      |       36|   0|   b|2805
          |   |   A:   A:    :       |      36|  Yes, that is cne of the things I like about it. |    |        |         |    |    |2806
          |   |   B:   B:    :       |      35|  Yes, but that isn't what I'm looking for. |    |        |         |    |    |2807
          |   |   C:   C:    :       |      24|  I don't know.                           |    |        |         |    |    |2808
          |   |   D:   D:    :       |      13|  No, but financial security is not what I want |    |        |         |    |    |2809
          |   |    :    :    :       |        |  from a job anyway.                      |    |        |         |    |    |2810
          |   |   E:   E:    :       |      12|  No, but I wouldn't want to stay very long |    |        |         |    |    |2811
          |   |    :    :    :       |        |  even if it did.                         |    |        |         |    |    |2812
          |   |   F:   F:    :       |      11|  No, this is one of the drawbacks of the job. |    |        |         |    |    |2813
```

DELETED TASKS:

Reformatting the 4% sample. The 4% sample, including a separate scanning of Answer Sheets A and C, afforded an excellent opportunity for assessing the reliability of the scanning equipment. Tests run on these records indicated close agreement between the two scannings, except on the highly speeded tests. The data from both scannings were compared to the original answer sheets to determine the source of any discrepancies, but errors were found in roughly equal number for both scannings. The errors resulted chiefly from the machine reading an answer where none of the options were marked. Efforts to use one scanning to flag errors in the other were therefore dropped.

The original plans called for a large number of new variables to be created from the item data on the 4% file. These plans were dropped for lack of funds (although the scoring programs were stored away for future use).

The retest tape. The 1963 retest data are unquestionably valuable for a thorough assessment of changes occurring during high school. Because the cases on the file do not constitute a true probability sample of high school students and because reformatting these data depended on the reformatting of the 1960 data, work on this file was scheduled after the main 1960 and follow-up files, and we never got to it. More will be done with these data as time and money can be scraped together to support such work.

The school data tapes. Work on the school data file was not included in the original grant for tape cleanup, but was included in the grant awarded a year later. An abbreviated school file was created for merging with the student data and considerable work was done in checking and correcting the weights assigned to each school and the test number ranges for each school. These data were merged with the Test Record Form data for each school. Furter work was suspended in order to complete the main tape documentation and revisions to the Data Bank Handbook.

With the completion of the recent grant for data tape reformatting and documentation, the TALENT data base is in far better shape than at any time previously. This is not to say that further improvements could not be made. Several editing tasks could not be carried out within the scope of the recent grant but are, nonetheless, important.

Chapter 14. COMPUTERS AND COMPUTER PROGRAMS

Everybody who uses computers has problems with them. Project TALENT is no exception. However, there have been bright spots too. This chapter is about both.

A. Problems with programs

We discovered early that most standard program packages were not well suited to use with large data sets, such as TALENT. When Bill Cooley joined the TALENT staff as Project Director in 1964, he brought with him the set of multivariate analysis programs developed in connection with his textbook (Cooley and Lohnes, 1962). It didn't take Bill long to decide that the programs would not work well on TALENT data. It was clear to him that the best approach would be for Project TALENT to develop its own data analysis programs. Therefore he added an expert programmer, statistician, and mathematician (all combined in one person) to the staff in 1965. That was Bary Wingersky. Bary wrote a very fine set of multivariate analysis programs and other data analysis programs as well. At that time we were using the Pitt facilities, which included two IBM computers-- a 7070, which we used for data processing, and a 7090, which we used for data analysis. Thus Bary's programs were designed for use on the 7090.

When we were preparing to move to Palo Alto, where we would be using the IBM 360/50 that AIR was planning to secure for the joint use of Project PLAN and Project TALENT, it was necessary to convert Bary's programs since the 360, it turned out, was not compatible with the 7090. Some of the programs got converted, but unfortunately a great many didn't. Bary left AIR at about the time we were switching from Pittsburgh to Palo Alto, and from a 7070 to a 360/50. Consequently in Palo Alto we lost the use of a great many of the programs that Bary and his staff had developed. Eventually Dick Williams found time to convert some of them, but most of them never were.

Thus in recent years (particularly since about 1970, when Dick Williams left) we seem to have become primarily dependent on program packages again.

Of course the packages have improved since the early years, but there are still many problems associated with their use.

B. Keeping up with technology

Over the years we have used many different computers. As Section A above may suggest some of the transitions (like the switch from the 7090 to the 360) have been traumatic events—even though in most of the transitions the newer computer (e.g., the 360 as a replacement for the 7090) was bigger, better, faster, and more versatile than whatever it replaced. Some of the transitions, on the other hand have been so uneventful that the change was hardly noticed.

As has already been indicated the computer that Pitt had when Project TALENT first needed access was a 650. That was soon replaced by 7070, which was designed primarily for data processing (as opposed to data analysis) and was excellent for the purpose. After a while, Pitt added a 7090, which unlike its stable-mate, the 7070, was intended primarily for data analysis, not data processing. Since the move to Palo Alto we have had the use of a series of 360's and 370's, and we have used the same computer for data processing and data analysis.

AIR no longer has its own computer, having found that rented time on a computer service agency's computer works much better in the long run. Currently we are using an IBM 370/165 and a 370/158.

Chapter 15. STRUGGLING WITH WEIGHTS
(Not a Chapter About Weight-Lifting)

Because not all schools in the sample were selected on the basis of the same sampling ratio, and therefore not all students in the sample had had an equal *a priori* chance of selection, it is necessary to weight sample members differentially if one wishes to eliminate systematic error, or at least reduce it substantially. However, in so doing one must realize that there is a trade-off. Every time differential weights are used one is increasing the random errors of the type usually called sampling errors. Minor variations in weights will have only a negligible effect on random error, but extreme variations can increase it seriously. Table 15-1, which classifies errors in a two-by-two table, provides a framework that may be helpful. (In this chapter we are concerned exclusively with the two top cells of the table — i.e., with sampling error, not measurement error.)

A. Sets of weights that have been developed

1. Weights for 1960 data (where no follow-up data are being used)

At least six separate sets of weights (Weights A, B, C, D, E, and F) have been developed for use in conjunction with the 1960 data. Weights B, C, and F are intended primarily for use in computing statistics in which the school is the unit of interest, while Weights A, D, and E apply to students.

Weights A and E are almost identical, differing only for students in the two junior high schools referred to in chapter 7 (on page 7-11 immediately above the "Author's Note"). Weight A for students in these schools is the same as the weight for the senior high school selected from the same school district. Weight E is 20 times Weight A for students in these two schools. Weight E gives as close to an unbiased estimate as it is possible to achieve with these data. But the weights for the two junior high schools are enormous, and for most purposes tend to inflate the systematic

TABLE 15-1. Taxonomy and etiology of statistical error*

		Causes of error or factors that increase the extent of error	
SOURCE OF ERROR	KIND OF ERROR	RANDOM ERROR	SYSTEMATIC ERROR (BIAS)
Using a sample instead of the entire population	Sampling error**	1. Not having a large sample 2. Selecting the sample in such a way that differential weights must be used. (The more extreme the differences in weights the greater the sampling error.)	a. Selecting the sample in such a way that not all population members have an *a priori* probability of selection greater than zero b. Not using weights that are the reciprocal of the *a priori* probability of selection*** c. Certain sampling techniques (e.g. cluster sampling) may cause systematic error even though the differential weights are mathematically correct****
Using imperfect measuring instruments	Measurement error	Using instrument that lack perfect reliability	Using instruments that lack perfect validity*****

* It is recognized that this table is to some degree an oversimplification.

** The term "sampling error" is usually reserved for random errors due to sampling. In this table, however, it is convenient to use the term, for the nonce, to include both random and systematic error.

*** In the case of multistage sampling, the probability of selection is the product of the various sampling ratios.

**** However with large samples systematic error from this source is likely to be negligible.

***** Bias in an instrument may be regarded as a special kind of invalidity.

errors far too much to be compensated for by the slight reduction in random error. Therefore Weight E is seldom used. Weight A is almost always preferable even though its failure to weight those two junior high schools precisely results in a little bit of systematic error. Random error with Weight A should be much less than with Weight E.

Weight D is the student weight used to get estimates based on the probability sample of 15-year-olds.

In regard to the weights for use in determining statistics about the schools, Weight B is identical with Weight A except for New York City and Chicago, in which, to compensate for the inclusion of extra schools, only a sample of the students in each school were tested. For schools in these two cities Weight B is much smaller than Weight A. (In New York City, where every secondary school was tested, Weight B equals 1.)

Weight B is 0 for schools that declined to participate in the testing. Weight C is the weight to be used when the statistics are to be based on all schools selected for the sample—not just those that agreed to participate.

Weight F equals Weight B except for the two schools for which Weight E exceeds Weight A; for those two schools Weight F is 20 times Weight B. As in the case of Weights A and E, Weight F will yield essentially unbiased statistics about schools, but with much larger random errors than Weight B.

When used in conjunction with follow-up weights, Weight A is usually referred to as "1960 Weight A".*

2. Weights for follow-up data

The follow-up weights differ somewhat from the 1960 weights because not everyone responds to the follow-up questionnaires.

---

* In mathematical formulas both the 1960 weights and follow-up weights are represented by W with a subscript representing time —"5" for 5-year follow-up, "11" for 11-year follow-up, etc.; and another subscript representing type of follow-up weight--e.g., "A" for follow-up weight A, etc. In this context the time code for 1960 is "0". Thus in contexts where five-year follow-up weight A is designated W_{5A}, 1960 Weight A would be designated W_0.

At least four separate sets of follow-up weights have been developed. They are Weights A, B_1, B_2, and C.

Follow-up Weight A is the same as 1960 Weight A, except for nonrespondents to the mailed questionnaire. For mailed-questionnaire nonrespondents not selected for the special sample follow-up, Weight A is 0. For special-sample members, the 1960 Weight A is multiplied by the reciprocal of the sampling ratio used in selecting the special sample. This applies whether or not the person cooperated in the special-sample survey.

Weight B_1 is the same as Weight A except for members of the special sample. Special-sample members who cooperated have their weights increased by a factor sufficient to compensate for those special-sample members who didn't. Weights for these latter members are reduced to 0.

Weight B_2, like Weight B_1, is 0 for special-sample members who don't cooperate. But to compensate for them, instead of just raising the weights of the other special-sample members, the weights of all respondents, whether to the mailed questionnaire or to the special-sample survey, are increased proportionately. Thus for cooperating members of the special sample Weight B_2 is less than Weight B_1 while the opposite is true of mailed-questionnaire respondents.

Follow-up Weight C* was developed in an effort to improve on follow-up Weights B_1 and B_2, by correcting in a more sophisticated way for noncooperating members of the special sample. Weight C has proved very effective. For most variables it gives far more accurate estimates than any of the other sets of follow-up weights we have tried. Like Weight B_2, Weight C distributes the weight for special-sample noncooperators among the individuals for whom questionnaire responses are available, whether via the mailed questionnaire or the special sample. Weight C is described in Appendix E2.

---

* Developed by the author of this History.

B. Whether and how to weight

With the multiplicity of weights described above, how does one know which one to use for a particular purpose? In fact, how does one even know whether to use any? Should one be using unweighted data instead?

There aren't any good rules of thumb for this last question. As has already been implied one has to consider the purpose of the research, and the trade-off between systematic and random error.

In general, however, weighting is essential when one is interested in normative or distributional data; in other words, when one wishes to be able to make statements of the following sort:

> The average spelling score for 12th grade students who later majored in English in college was __________.
>
> The 84th percentile for 9th-grade girls on the Scientific Attitude Scale in 1960 was __________.
>
> Eleven years after high school the average salary of members of the 10th-grade cohort who had become teachers was __________.

Generally speaking, for data analyses of this type, 1960 Weight A would be the preferred weight if the entire 1960 sample for the segment of interest (e.g., 10th-grade girls) or a representative subsample of it is being used. If the analysis is limited to cases with follow-up data, follow-up Weight C is usually the best one to use (at least in the opinion of this author). Note that 1960 Weight A is to be preferred to 1960 Weight E even though Weight A does not yield quite as "pure" results (i.e., results free from systematic error). That minor disadvantage is more than cancelled by the reduction in random sample errors.

But if one is interested in being able to make statements of a relational or predictive rather than normative type, it is the opinion of this author* that on the whole one is better off not using weights at all. Weighting the cases differentially will not have much effect on correlation coefficients except to make them less stable--in other words, to increase the random error. And if a statement of the following type,

> "Students with high math scores but low scores on Mechanical Information and Mechanical-Technical Interest are more likely to be successful and happy in __(Occupation A)__ than in __(Occupation B)__."

were to fall to pieces just because it was based on a sample that wasn't perfectly representative of the national population, it wouldn't be worth much to begin with.

C. Self-weighting: A solution to the trade-off dilemma

In Section B above, there was some discussion of the trade-off between not weighting--in other words, using unit weights--and using differential weights. Unit weighting has the advantage of reducing sampling errors but it may introduce significant systematic errors. The solution to the problem was initiated by John Flanagan when he proposed that we develop "self-weighted subsamples" of the TALENT sample. These would be subsamples selected in such a way that each case in the TALENT sample would have a probability of selection proportional to its 1960 Weight A. (For a self-weighted sample of the cases with follow-up data, presumably follow-up Weight C would be used rather than Weight A.) The advantage of such a subsample would be that the optimum weights would be unit weights. In other words, the subsample would be "self-weighted."

Marion Shaycoft developed an algorithm for selecting a subsample that would meet this requirement and that except for the differential-probabilities feature would be a simple random sample. Don McLaughlin

---

* I feel obliged to point out that not everyone agrees. Some very knowledgeable people have a contrary opinion on this point.

--Author

proposed an alternative algorithm that would yield a systematic (rather than simple random) subsample having the self-weighting feature. It was unanimously agreed (by Flanagan, McLaughlin, and Shaycoft) that Don's approach would have major advantages. Consequently it has been used frequently since then. It was, incidentally, the procedure John Flanagan would later use in selecting a 1000-case sample from the probability sample of 15-year-olds to interview 15 years later for a quality-of-life study of 30-year-olds. (See Chapter 21, Section E)

The same approach was utilized in a slightly modified form (i.e., with stratification) to select the special sample of nonrespondents to the 11-year follow-up of the 9th-grade cohort.

Chapter 16. OTHER STATISTICAL PROBLEMS ENCOUNTERED AND SOLVED

The title of this chapter is not meant to imply that we have solved every statistical problem that we have encountered. What it means, instead, is that we are judiciously limiting this chapter to the problems we have solved. In accordance with the HIBK syndrome (see Preface) the solution to a few of the problems, unfortunately, came a bit too late to use in the present project. Ah, but in Son-of-TALENT. . . .

Since we were discussing sampling in the last chapter, we shall start with that topic.

A. Sampling

1. How to handle schools that are in the project on a specially invited or volunteer basis

As discussed in Chapter 7 Section B, quite a few schools participated in the project that hadn't been in the initially selected sample. Some were invited by us (for instance, all the schools in Knoxville that hadn't originally been selected) and some invited themselves. We treated the students in all these schools as not being members of the probability sample. This was technically sound but *was* it really necessary to lose all the good data from these schools when doing analyses requiring a probability sample? The answer is that a method was developed by the author of this History, a couple of years ago, in connection with an activity unrelated to Project TALENT, whereby it would have been possible to treat any of these "non-probability-sample schools" that had been in the original universe (i.e., all secondary schools in the United States) as members of the probability sample, through appropriate adjusting of the weights. The method is described in Appendix D1.

2. The "feeder junior high school" problem

As discussed in Chapter 7 Section B (page 7-9), a special procedure has now been developed to cope with the problem of selecting the "feeder" junior high schools for the sample. The procedure, which was developed far too recently to have been available for use in selecting the TALENT sample in 1959, is called the "closed school-cluster procedure." It is described in Appendix D2.

3. Coping with complexly structured samples in computing standard errors

Although this problem is concerned primarily with computation, it is put under the sampling heading because it has implications for sampling. (It involves picking two wholly parallel subsamples.) Like the procedures discussed in paragraphs 1 and 2 above, this one has not been tried yet. It is described in Appendix D3.

4. Is the Special Sample of nonrespondents necessary? (An empirical validation of follow-up Weight C)

Collecting data from the special sample of nonrespondents is extremely costly. Therefore there has been serious consideration at one time or another of whether we really needed it; there have also been at least two empirical studies.

The first was an APA paper presented by Marion Shaycoft in 1964 (reprinted as Appendix E1), called "When is Bias Better?" This study was an effort to determine whether by stratifying on 1960 data and weighting the questionnaire respondents in each stratum to also represent the nonrespondents it would be possible to dispense with the special sample. The answer was a qualified "no."*

---

* It was not an unqualified "no," because the study was somewhat restricted by the fact that the range of computer programs available to us was very limited and did not permit stratification on more than one variable.

The second empirical study turned out to be a validation of follow-up Weight C (described briefly in Chapter 15 Section A2). This was a study by Don McLaughlin and Wendy Yen, in which statistics obtained by using Weight C were compared with corresponding statistics computed using respondents only, but adjusting their weights. Weight C turned out to be superior to any of the other procedures used. The study is reported in "Project TALENT's Special Sample: Is It Necessary?" (McLaughlin, Fulscher, and Yen, 1974).

5. Can the selection of the Special Sample be improved?

As indicated in Section 4 above, empirical evidence suggested that use of a special sample should be continued. However a substantial improvement has been introduced recently by using a modified self-weighting scheme with stratification on variables of particular interest and judicious oversampling of certain cells.

B. Reducing sampling error--both random and systematic

As has been implied at several points, use of a special sample of nonrespondents has tended to increase systematic error, though it reduces random error. Efforts have been more or less continuous to find ways to moderate the increase in random error and thus decrease in total error.

The development of Weight C (see Chapter 15, section A2) was one such effort. Use of a self-weighted special sample (as in the 11-year follow-up of the grade 9 cohort) was another.

A third was initiated in 1975, when Laurie Wise and Don McLaughlin, in an in-house memo, proposed a formula for getting a weighted mean of data from mailed-questionnaire respondents and special-sample respondents, with the weights for the two components being different from those which, like follow-up Weight C, are intended to get as close to an unbiased estimate as possible. The goal of the McLaughlin-Wise procedure was to minimize total error. But their formula for determining the weights for the two components was a preliminary one, requiring many assumptions, some of which seemed dubious.

Starting with their idea, the author of this History developed a set of formulas and procedures, applicable under various conditions and not (we hope) requiring untenable assumptions. The outcome was a paper entitled "Minimizing Total Error: A Trade-Off Problem" (Shaycoft, Wise, & McLaughlin, 1975), which is reproduced as Appendix E3.

C. Reliability coefficients and related statistics

1. Reliability coefficients

Parallel form reliability coefficients were not feasible for the TALENT tests because there were no parallel forms. Retest correlations are unsuitable for use as reliability coefficients for cognitive variables. That meant we were limited to some sort of internal consistency reliability coefficient--either split-half or one of the KR coefficients. It is our judgment that for most of the tests in the TALENT battery, split-half coefficients with *a priori* splits are preferable to KR-20 or KR-21. In the first place K-R coefficients would tend to systematically overestimate the reliability of certain tests such as Reading Comprehension for which split-half coefficients with an appropriate item split would not. Furthermore, formulas such as KR-20 are based on the assumption that all the items are fungible units. That assumption may be valid for certain scales in the battery, such as Visualization in Two Dimensions (for which the reliability cannot be determined anyhow, because it is a somewhat speeded test and therefore a parallel form or separate administrations would be needed) but it certainly isn't true for the information scales and achievement tests, for which, it will be recalled from Chapter 5, very detailed rationales were prepared in advance, which spelled out the proposed distribution of item content. Therefore the method of choice for determining the reliability of the TALENT tests is split-half, using *a priori* splits where appropriate, with Angoff formula 16 (Angoff, 1953) as the correction formula. This has limited our data sources to the 4% sample (see Chapter 12) which provides suitable data for all variables except, of course, those that are substantially speeded and those few for which there is some other factor precluding reliability determination (see Shaycoft, 1967, Chapter 4, pp. 4-12 through 4-15) or to the 1963 retest data (see Chapter 11) for the variables for which grade 12 data are available.

2. Coefficients of internal precision

Around 1968, Lyle Schoenfeldt was working on a study of twins in the TALENT sample. One day he and I were discussing some strange results he was getting, which neither he nor I understood. When he tried to apply a conventional correction for attenuation to the correlation between twins' scores he got suspiciously high correlations; some of them were over 1. Such results are possible only when one is using as the divisor an underestimate of the geometric mean of the two reliability coefficients. But we had no reason to think that our reliability coefficients were systematic underestimates. The explanation suddenly occurred to me, and a lot of other things fell into place. The explanation lay in the fact that when the correlation is between two measures obtained with the identical test the appropriate coefficient to use in correction for attenuation is not the reliability coefficient, because using the reliability coefficient means that a correction is being applied for, among other things, the random effects of having a different sample of items. This is an overcorrection when one has the same items in the two variables correlated. In a Eureka-type of reaction, it occurred to me that the same explanation would account for some inexplicably high corrected-for-attenuation correlations, many of them far in excess of 1, that I had obtained in the retest study (Shaycoft, 1967, Table 5-4, columns 3 and 4), and which I had found baffling at the time. The correlations were between the grade 9 and grade 12 scores of the same student on the same test.

The solution to the problem of these excessively high "corrected" correlations--not only the ones Lyle had obtained in his research on twins and those I had obtained in research involving retesting with the same test, but probably others as well--was to derive a formula for a coefficient that would be like a reliability coefficient except that it would not include the concept of sampling errors in selecting the particular item sample as a source of error. Hence, the coefficient of internal precision! (See Appendix C.) These coefficients have been computed for a limited number of tests in the TALENT battery, but eventually they should be computed once and for all, for all of the tests for which they are suitable. This would make them available to any researcher using the TALENT data, since the same correction-for-range formula applies as for reliability coefficients, thus making it possible to adapt the "once and for all" coefficients to the particular sample of students being used.

D. Inferring score distributions for an unselected age group

In 1963 we had a contract with the Air Force, which called for inferring score distributions for the total population of 18-year-olds in the United States. The solution, which is presented in Appendix F, would presumably be generalizable to other age groups (such as age 17).

E. Problems in methodology related to career guidance

Appendix G consists of papers by the author of this History on two problems of a methodological nature, to which approaches that are believed to be largely new were developed in the course of Project TALENT.

The first is a new approach to the use of a battery such as TALENT in career guidance; it is called propinquity analysis. See Appendix G1; also for further information, see Chapter 7 and Appendix N of Five Years After High School (Flanagan, Shaycoft, Richards, and Claudy, 1971).

The second is a suggestion on the grouping of career categories. (See Appendix G2). Attempts at hierarchical analysis have led to the conclusion that it isn't a particularly useful approach for the TALENT data. (Flanagan, Shaycoft, Richards, and Claudy, 1971, Chapter 5.)

Chapter 17. MANAGEMENT PROBLEMS

Whatever problems are routine in almost any research management situation are greatly magnified in a longitudinal study. And if it is a large-scale longitudinal study, the problems multiply.

This chapter is concerned not with management problems in general but with those problems that apply particularly to the management of a large-scale longitudinal study.

A. Record-keeping problems

1. Identification numbers and other codes

In a longitudinal study it is completely impractical to rely on students' names or school names or almost anything else that is spelled out in words, if one wishes to maintain continuity--for instance to be able to match new follow-up data with the corresponding set of earlier data. Some sort of identification code is necessary--either a serial number or some kind of segmented number where each segment has a meaning, or perhaps a code that is a combination of numbers and letters. (For computer use all-number codes are usually somewhat preferable to number-and-letter combinations.)

2. Some comments about redundancy

Whatever the code system is, one has to bear in mind the fact that everyone makes errors sometimes and some people make errors very frequently, so that matching records on the basis of identification number is not error-proof. Someone may have made an error in recording a number or in keypunching or keytaping the number--or if the number is position-coded on a machine-scannable form, the scanner may fail to pick it up completely. The same caveats apply not just to identification numbers but to other important categorizing or identifying information.

Therefore it is important to build enough redundancy into the system to make it possible to cope with error--both human error and machine error.

In Project TALENT lots of redundancy was built into the system. In addition to the six-digit identification number, each answer sheet contained either the student's full name, or his last name and his initials, or, at the very least, the first several letters of his last name. The student's sex was position-coded on two of the five answer sheets, as was his grade. In addition each set of answer sheets was preceded by a "Group Identification Sheet" filled out by the teacher or other examiner, and containing in position-coded form the grade as well as other useful information. The answer sheet form containing the student's full name, address, and other basic information contains in position-coded form not only his date of birth (month, day, year) but also his age in years at his last birthday. This particular bit of redundancy seemed essential even though we made it impossible for these spring 1960 examinees to record their year of birth as 1960. (We prevented that particular error by not providing a position for position-coding '60.)

But it seems that the more carefully one builds redundancy into the system, the more likely it is that some well-meaning person will come along and wipe it out before anyone gets a chance to take advantage of it (see Chapter 13). He will wipe it out because somewhere along the way he has been taught that redundancy is a bad thing.

Therefore our advice to future longitudinal researchers who want to profit from our experience is that by all means they should build plenty of redundancy into the most important parts of their record-keeping systems, but that they should also make certain that anyone who is in a position to wipe out the redundancy prematurely is informed that it is there, and is not there by accident, and they should also make certain that he understands what it is there for. (This underlines the importance of good two-way communication between research staff and computer staff and the undesirability of having them located at different sites; see section E2 in this chapter.)

3. Confidentiality

In releasing research results or individual data of any kind, it is important to maintain commitments on confidentiality and not infringe on sample members' rights of privacy. This requires vigilance. Routine deletion of names and dates of birth from released information is not sufficient. It is necessary to make certain that identity cannot be inferred through the use of mathematical logic. For instance in one system of classifying states into regions the Far West might consist of Washington, Oregon, and California, while in another system the Far West would be defined as Washington, Oregon, California, and Nevada. If the same data were released by region both ways, it would be easy to determine the values for Nevada by subtraction. Researchers and research administrators have to be alert to this sort of possibility, and keep it from happening.

4. Retrievability

In a longitudinal study the need sometimes arises to correct an error in an individual record or to update the record without waiting until there are a hundred other records to correct or update at the same time. Or the need arises to get a printout of the record for a single case. A data processing system that permits handling such situations efficiently and economically is worth developing.

B. Securing and maintaining cooperation

Project TALENT has developed certain procedures and techniques to help keep members of the sample interested and cooperative. These include such devices as an ID card identifying the person as a member of the TALENT sample (see Chapter 7 section C); and the Project TALENT News, a 4-page newspaper mailed to the entire sample once a year (see Chapter 18, section E1).

In order to be able to keep sample members interested and cooperative, one has to know how to contact them. In other words, up-to-date name-and-address files are essential. One thing this means is that about once a year (or at least once every year-and-a-half) something--a postcard, a form letter (preferably, though, one that doesn't look too much like a form letter), a newsletter or newspaper, a questionnaire, something--should be mailed to everyone in the sample, with "Address Correction Requested" on the envelope. That is an excellent way to help keep address files up-to-date. The reason it should be done once every year or year-and-a-half is that postal regulation require post offices to keep address changes on file for two years; after that they throw them away. It is worthwhile to assign someone on the staff responsibility for keeping informed about changes in postal regulations. There is nothing immutable about those regulations--and changes one is unaware of or changes one finds out about too late can raise the project's mailing costs by thousands of dollars, or can result in having to drop thousands of entries from the project's mailer tape because they have turned into dead-end addresses.

It is also worthwhile to do a little experimenting on what mailing "styles" are most likely to result in cooperation. For instance, if one is mailing a questionnaire first-class, does a postage stamp have any advantage over metered postage? Does a commemorative stamp have any advantage over an ordinary postage stamp? Some evidence was obtained years ago that the answer to both of these questions was affirmative. Whether it still is isn't entirely clear, since collective moods change with the times and so do prevalent mailing procedures. (Metering is used much more frequently for first-class mail than it used to be, so the sight of an envelope with metered postage stamped on it isn't as likely to stimulate the mental response "junk mail" or the reflex action of tossing the thing in the circular file.) As a matter of fact, the entire problem of when it pays to mail something first-class that could legally go by bulk rate has to be considered. As postal rates keep rising it becomes more and more important to send mail by the most economical way that will accomplish whatever you are trying to accomplish through that particular mailing. And it is probably not feasible to develop a "once-and-for-all" answer, since not only absolute costs but the relative costs of first-class and bulk-rate mailing keep changing.

Furthermore before sending something out by first-class mail that could go by bulk-rate, you should know just what it is that you are paying extra for. Are you buying the free forwarding? a favorable impression on the recipient? more careful handling by the post office? faster delivery? Whatever benefits you expect first-class mail to deliver, it is important to be aware of your expectations so that you can take steps to find out whether they were fulfilled, and perhaps learn from experience.

Closely related to the question of how the researcher mails things out is how he plans to get them back. Should he enclose a self-addressed "business reply mail" envelope with a first class mailing permit? Or should the return envelope have a commemorative postage stamp affixed to it? The mailing permit means there will be no charge for return postage in the case of nonrespondents--but for each respondent it will cost more than a postage stamp would have. And furthermore, there is evidence that having a stamp already attached to a return envelope increases the likelihood of response. (Of course the extent to which this is true must depend, to at least some extent, on the nature of the group on which one is doing the longitudinal study.)

The next chapter contains some further discussion of Project TALENT's mailing procedures for follow-up questionnaires. The reader is also referred to "Methodology of the Project TALENT 11-Year Follow-Up Study," (Rossi, Wise, Williams, and Carrel, 1976.)

C. Getting funding (the cliff-hanger serial)

In Chapter 2 (Sections B and D in particular) the problems of getting funding to get the Project started were discussed. In April 1959 (just a little more than two years after the idea first began to take shape) a $480,000 grant was received, to get the project going. But obviously that $480,000 was not going to put a permanent end to the funding problem. On a project that it was anticipated wouldn't end until, perhaps, about 1985 (2 years after the 20-year follow-up of the high school class of 1963, in other words the grade 9 cohort), it was clear that getting funding would be a recurrent problem--indeed, a nearly continuous problem. At various times the project's primary funding agency (OE initially, then NIE) has renewed the funding, sometimes with a five-year grant, sometimes with a two or two-and-a-half year grant, sometimes with just a one-year grant. (What the length of the grant is probably depends more on what the fiscal situation, the prevailing administrative regulations, and the general mood are at the funding agency at grant-renewal time than it does on any specific formal steps that AIR can take (such as writing a 5-year proposal versus a one-year proposal).

Of course the best insurance on getting the funding continued is doing as good a job as possible on the study--but of course even that provides no sure guarantee against loss of funding.

One way or another, the project has had continuous funding since 1957--but on several occasions there have been delays of many months in our receiving the signed document, so that we have been forced to mark time, rather than going full speed ahead with our plans.

The most troublesome delay occurred when we were completing the 5-year follow-up and were seeking funding for the 10-year follow-up, which should have started in 1970. At that point it wasn't clear where our funding was going to come from. The feeling in certain quarters at OE was that Project TALENT should certainly be continued, but since the sample members would then be about 28 years old, and thus well out of the age range that was of most concern to the Office of Education, it wasn't clear that OE should continue to be the prime funding agency. What about the Department of Labor? How about the Department of Defense? Any why not National Science Foundation? (We duly explored all these leads.) The part of OE that had in recent years been providing the support for Project TALENT was the National Center for Educational Research and Development (NCERD). Most of the people we had dealings with at NCERD, including the monitors (Sue Klein and then Larry Goebel) understood the need for continuing the project and were strong supporters. But since money was very tight at NCERD it was suggested that even if OE were to provide the main support for the project perhaps the funding could be a joint venture of several parts of OE, rather than coming exclusively from NCERD. It was suggested that we explore this possibility with several other parts of OE--for instance the Bureau of Education for the Handicapped (BEH), the Bureau of Elementary and Secondary Education (BESE), the Bureau of Higher Education (BHE), the Bureau of Educational Personnel Development (BEPD), and the Bureau of Adult Vocational and Technical Education (BAVTE).

We conferred with representatives (usually the Director) of all these groups and submitted special "sub-proposals" showing how Project TALENT data could fit in with their particular needs and interests. The persons with whom we conferred were all quite interested, and they recognized how Project TALENT data might be useful to their agencies; but unfortunately most of them had no unbudgeted discretionary funds available. The monitor (Larry Goebel) also explored with OE agencies other than NCERD the possibility that they might provide supplementary support for Project TALENT. Unfortunately his efforts were no more successful than ours. It soon became apparent to us that if OE was going to provide support for the 10-year follow-up, the part of OE that was going to provide that support would have to be NCERD. That became apparent to NCERD, too. But since the fiscal situation was too tight to enable them to manage it with funds available in FY '71 (the fiscal year running from July 1970 through June 1971) it was necessary to wait until the FY '72 funds became available. And that is why we had an 11-year follow-up instead of a 10-year follow-up! In the intervening year it was possible to keep the project in existence, and thus maintain staff continuity, via time extensions on the previous grant, from which unexpended funds still remained.

At just about the time that NCERD was arranging to fund the 11-year follow-up, one of our efforts to get supplementary funding from another agency finally proved successful; we received word that NSF was going to supplement the OE grant (approximately $1,300,000 for the first 30 months) with a very welcome additional $75,000. (NSF arranged, via an interagency agreement, to have their grant administered by OE.)

The next major episode of uncertainty and delay (though not as long a delay) came when the grant for the first 30 months of the 11-year follow-up was about to expire. By then NIE was in existence and Project TALENT had been transferred to it from OE. NIE had appointed a panel on longitudinal research studies and methodology. One of the functions of this panel was to serve as a Review Panel for Project TALENT's proposal for continued funding. A meeting was held in Washington on 18 October 1973, at which TALENT staff members presented reports, and the committee made recommendations. The meeting was attended by:

Panel Members

- Jerry Bachman
- Robert Crain
- Christopher Jencks
- Ned Gramlick
- Herbert Parnes
- Michael Timpane

NIE staff

- Garry McDaniels (chairman)
- Larry Goebel (TALENT monitor)
- Andrew Porter
- Marshall Smith

TALENT staff

- John Flanagan
- Bill Clemans*
- John Claudy
- Don McLaughlin
- Marion Shaycoft

The Panel made several recommendations. The most important one (at least from our viewpoint) was that funding be continued; however they recommended that it be earmarked primarily for data collection and processing, with only a comparatively small amount being allocated for data analysis. Most of the rest of their recommendations reflected concerns about the data tapes, with reference particularly to their organization, documentation, and availability--in other words many of the same concerns that a few of the staff members had already had and that are touched on in Chapter 13.

---

* Bill, at that time, had agreed to become TALENT's next Project Director, but he had not yet joined the staff.

Specifically they recommended that data tapes should be made available to researchers on request, and that AIR prepare a proposal for revamping the tapes and preparing thorough documentation. (They suggested that a full-time project director and two programmers could complete this work in a year--an entirely reasonable estimate, in the opinion of the author of this History.)

About a month after the October 18 meeting Marion Shaycoft was pleased to receive a request from the TALENT monitor, Larry Goebel, to draft a proposal for getting the tapes and their documentation in good shape. She did this immediately, incorporating in the proposal the recommendation that a new Data Bank handbook be prepared--along with recommendations for a comprehensive overhauling of the tapes and the tape-format documentation. (This proposal was finally funded about a year-and-a-half later. For details about what happened next, go back to Chapter 13.)

A grant was received for a year's work starting October 1976, to develop plans and instruments for a 17-year follow-up. That work is now in progress, and hopes are high that funds will be received for implementing the plans--in other words for actually carrying out the follow-up. Tune in next week. . . .

Total funding received thus far for "Mainstream Project TALENT" (i.e., the basic funding to keep the project going) is a little over 7 million dollars. Other TALENT-related studies done by AIR (exclusive of "Data Bank studies") total a little over a million dollars, bringing the combined total to something over 8 million dollars. Table 17-1 shows the breakdown.

D. Problems related to staffing

In addition to the problems touched on in this section the problems discussed in Section E, which follows, also have some implications for personnel management.

TABLE 17-1. TALENT Funding* for first 20 years (1957-1976)**

Funding agency	"Mainstream TALENT"	Other TALENT-related studies***	Total
OE	$ 4,833,000	$ 30,000	$ 4,863,000
NIE	1,920,000	211,000	2,131,000
NIMH	116,000	148,000	264,000
NIAAA	-	33,000	33,000
NICHD	137,000	250,000	387,000
Welfare Office****	-	67,000	67,000
NSF	90,000	166,000	256,000
ONR	35,000	31,000	66,000
AFPTRC	-	94,000	94,000
ARI	-	96,000	96,000
Other (approx.)	-	100,000	100,000
TOTAL	$ 7,131,000	$ 1,226,000	$ 8,357,000

---

* To the nearest thousand dollars

** As of December 1976

*** Excluding "Data Bank studies". (See Chapter 22.)

**** Part of HEW

1. Turnover

Obviously the longer a study lasts the greater the turnover and the problems associated with turnover. Project TALENT has had a lot of turnover as can be seen from a perusal of Chapter 8--and particularly from a look at the number of columns in Figure 8-2. But the fact that the Principal Investigator has been a constant has helped to keep to a minimum the problems that having half a dozen Project Directors would otherwise have caused.

2. Personnel selection

In this section I do not intend to talk about the importance of careful selection of staff in general because although that is obviously an important matter it is no more important for a longitudinal study than for any other kind of research.

Instead I shall focus on the lowly coding clerk--because there has sometimes been a tendency to underestimate the value of their work and its long-lasting effects. A researcher who wishes to do a study comparing career plans one year after high school with the same groups' career plans 11 years after high school is dependent for the one-year follow-up career plans on the work done by coders 12 to 15 years ago; and for the 11-year follow-up codes, on the work of coders 2 to 5 years ago. The two sets of coders are obviously different sets of people (that is the nature of the job), but it is desirable to be able to have confidence that they followed the same general principles--and that they did their job well. The coding is not a strictly routine job; doing it right requires intelligence and judgment, as well as accuracy. Moreover as the nature of the work load shifts from month to month, at some stages calling for questionnaire check-in, then coding, and then, months later, perhaps telephone interviewing of nonrespondents, obviously members of the temporary "coding staff" have to have the flexibility to cope with these shifts in assignment.

Therefore the more attention is paid to ability in selecting the coders the better the results are. At times there has been a tendency among those who do the hiring, and among the immediate and not-so-immediate supervisors of those who do the hiring, to overlook this. There are very large individual differences in applicants in terms of the specific abilities required; and these differences cannot be determined by looking at an application blank or conducting an interview.

E. Management problems related to computer and computer-related operations

1. Management of keypunching operations

The author of this History takes a dim view of large-scale keypunching operations if there is any effective and economical way of bypassing them. But sometimes there isn't. For instance in the one-year and five-year follow-ups, we had the enormous task of keypunching all questionnaire responses. (As is explained in Chapter 18, Section B 1, by the time of the 11-year follow-up we were able to switch to an optically scannable questionnaire form, but this option wasn't available to us on any practical basis at the time of the one-year and five-year follow-ups.)

If one can't avoid undertaking a large keypunching task the next question concerns who should do it. Should the organization (e.g., AIR) have its own keypunches and its own keypunching staff, or should the task be farmed out, to an agency whose primary function is keypunching? We have done it both ways. When Project TALENT was in Washington we had our own keypunches and keypunch operators. (That was during the one-year follow-up of the grade 12, grade 11, and grade 10 cohorts.) In Pittsburgh, and subsequently in Palo Alto, on the other hand, we farmed the keypunching out (to a different agency almost every year).

Most service agencies seem unwilling to guarantee that the error rate will be smaller than 1%, in terms of cards containing errors. The rate is much lower in terms of percentage of key-strokes that are wrong, but this is irrelevant since a single keystroke error in certain positions can make the entire card (or tape record) worthless.

There is no doubt in the author's mind that doing the keypunching right on the premises is by far the better procedure. One is able to set, and maintain, much higher standards of accuracy when the operation is under one's control than most of the outside agencies even attempt to meet (regardless of their claims). The costs of the two approaches are roughly comparable, and the greater accuracy of the cards produced in an in-house operation results in savings in subsequent processing--to say nothing of the research advantages of having significantly more accurate data.

The purpose of this discussion has been to give future researchers who have to decide whether to farm out keypunching or do it themselves the benefit of Project TALENT's experience (or at least the author's interpretation of that experience).

2. Coordinating between the research staff and the computer staff

As anyone who has read Chapter 13 will probably realize, a great many (though by no means all) of Project TALENT's problems with the data tapes had their start when there were 250 miles separating the computer staff (in Pittsburgh) from the rest of the project staff (in Washington). In a study such as Project TALENT, good two-way communication between the research staff and the computer staff is of primary importance. It is also desirable that the programmers, and particularly the programming director, be recognized as central members of the project staff rather than just being on the fringe of it. (What the situation has been in Project TALENT in this respect has varied from time to time, and the results have varied accordingly.) And as has been mentioned in previous chapters the computer and computer staff should not be 250 miles away!

3. Backup tapes

Having a duplicate set of tapes to serve as backup in case anything happens to the primary set is absolutely crucial in a project such as TALENT. Without such backup major subsets of the TALENT data would have been accidentally, but permanently, destroyed years ago.

F. Pitt leaves TALENT before TALENT leaves Pittsburgh!

It will be recalled from Chapter 2 that because a Federal law limited OE grants under the Cooperative Research Program to universities and state education departments the Project TALENT grant went to the University of Pittsburgh, not to AIR. Although this arrangement had advantages for the project (e.g., free computer time at first, then computer time at extremely low rates) it also created problems (for the project staff, and probably for the University's administrative staff as well). Therefore we were all pleased when the law was changed in such a way that the Project TALENT grant could be transferred to AIR. The three concerned parties--OE, Pitt, and AIR--jointly agreed to the transfer, which was to be effective January 1, 1967. The details were spelled out in a document called a "Novation Agreement", which represented a tri-partite agreement (Pitt, AIR, and the Unites States of America). In return for turning over administration of the project to AIR, Pitt requested, and received from AIR, a commitment that for the duration of the grant then in effect, any reports produced by Project TALENT would indicate that they represented the joint efforts of the American Institutes for Research and the University of Pittsburgh. And that is why our report entitled "Five Years after High School," (Flanagan, Shaycoft, Richards, and Claudy, 1971) bears on its cover the words "American Institutes for Research and University of Pittsburgh."

Chapter 18. THE FOLLOW-UPS

The original plan called for follow-ups one, five, ten, and twenty years after the grade cohort was scheduled to graduate from high school. In this chapter we shall talk about the procedures that have been used to execute this plan (and about changes that have been made in the plan itself).

A. Mailing of the questionnaires

Basically our follow-up approach has been to use a mailed questionnaire. The procedure has involved four "mailing waves." The first wave goes out in the early fall, typically at the end of September or the beginning of October; then about a week later a reminder postcard is mailed; then, about 2 weeks after that, the second wave of questionnaires is mailed to those from whom responses have not been received from the first wave. Then about 7 weeks later the third wave is sent to those who still haven't responded; the fourth wave is sent after another 2 months or more, to the "hard-core nonrespondents". Questionnaires in the four waves are identical, except that they carry an inconspicuous code indicating which wave they are, and each wave is usually printed in a different color; also if we are using an optically scannable form (as we have been doing on everything subsequent to the five-year follow-up) each questionnaire carries a set of machine-readable "skunk marks"--a series of broad black bands separated by white space--to designate the mailing wave.

Each questionnaire was accompanied by a letter--with different letters for different wave.

Further details about the mailing are given in various TALENT reports (e.g., Rossi, Wise, Williams, & Carrel, 1976, Chapter 2; Flanagan, Shaycoft, Richards, & Claudy, 1971, Chapter 1; Flanagan and Cooley, 1966, Chapter 2).

B. Processing the returned questionnaires

1. Keypunching vs. optical scanning

Our early follow-ups (one-year and five-year) were not on optically scannable forms. This meant that instead of having the completed questionnaire forms run through an optical-scanning machine that would convert the responses (and position-coded identifying information) directly to magnetic tape, every questionnaire had to be keypunched. Some years three IBM cards were required per questionnaire, and some years four. The first card was always a check-in card, which was used to identify respondents to earlier waves, so that their names could be eliminated from the mailings of subsequent waves. Subsequent cards contained item responses.

Those who have read Chapter 7 Section D can readily (and correctly) infer that the author of this History takes a dim view of any large-scale keypunching operation if there is any reasonable way of bypassing it in getting from document to magnetic tape. Thus it is probably obvious to the reader that if in each of eight follow-ups (one-year and five-year, four grades each) we undertook the horrendous task of getting 100,000 or more cards keypunched and verified, there was, in our judgment, no good way of avoiding it. We were continually checking on what the new technological developments in optical scanning were, in order to decide whether the time had at last come that we could switch from a questionnaire form that required keypunching to one from which output tape could be produced directly. It wasn't that good optical scanning machines didn't exist. They did exist, and they were getting better all the time. But during our earlier follow-ups the kinds of forms they could handle were not suitable for our needs. They were usually too restricted in size to allow room for all the questions we needed in the questionnaire, and too restricted in allowable format to give us the flexibility we needed.

By about the time we were getting ready to plan the 10-year follow-up, however, National Computer Systems (NCS), in Minneapolis, had developed optical scanning equipment, document printing procedures,and preliminary processing routines that would meet our needs. Using their services would enable us to get our data tapes prepared more economically and with substantially better accuracy than we could achieve via the keypunching route. Furthermore the NCS machines could handle questionnaires folded accordion-style, with as many as six 7" x 11" sheets (12 pages). This would actually give us more space for questions than the keypunch-style documents, which were typically 8-page 8½" x 11" booklets with the letter occupying the front page and the mailing address and return address the back page, would allow.

Taking everything into consideration, we decided to use NCS's services starting with the 10-year follow-up. As has already been noted, however, there never was a 10-year follow-up--so we didn't initiate use of an optically scannable questionnaire until the 11-year follow-up.

2. Coding

Regardless of whether the questionnaire was to be keypunched or optically scanned, some preliminary processing of the returned questionnaires, at AIR, was necessary. For certain questions (e.g., name of college attended, kind of job, long-range career plan) a multiple-choice item format on the questionnaire was not feasible; the respondent had to write the answers to such questions. This meant that the responses had to be converted to numerical codes, and in the case of optically scannable questionnaires the code then had to be recorded on the questionnaire form by blackening appropriate circles on a grid, to produce a "position code." This is done by one coding clerk, and checked independently by another; the coding supervisor, of course, is responsible for spotchecking as much as necessary, and making certain that appropriate standards of accuracy are met. Coding is a seasonal occupation, and it is performed by part-time clerks but it is by no means a strictly routine job. Doing it well requires speed and accuracy--but also intelligence and judgment. (Most of our coding clerks are college students who want part-time jobs; quite a few have eventually become members of AIR's permanent research staff, and are doing research studies themselves.)

3. Name and address changes

In addition to the coding of responses to open-end questions, an important part of the coding clerk's job is to prepare rosters for changes in a sample member's name or address or both. Correction cards are then keypunched from the rosters, so that the mailer tape can be corrected.

C. Mail response rates

On the very first follow-up survey we did--the one-year follow-up of the grade 12 cohort--almost two-thirds of the recipients completed the mailed questionnaire. Actually not everyone who should have been mailed the questionnaire was, since our mailer tape (the tape with names and addresses) wasn't in as good condition as it should have been, so that the response rate to the mailed questionnaire was only about 62 percent of the total 12th grade sample. That rate, though,--the "only 62%" rate--was really extraordinarily good, in view of the fact that response was purely voluntary. Never again have we even come close to a 62% response rate. At the time we achieved it, we thought it was a very good response rate; now we think it was sensational! (Of course, the fact that for this follow-up the least time had elapsed helped. Not only was the Project TALENT experience freshest in the sample members' minds, but our address records were freshest too.

For the one-year follow-ups of the other three grade cohorts, the response rates were in the 40's. The overall mailed questionnaire response rate for the one-year follow-up was about 50 percent.

For the five-year follow-up, grade 12 was still the one on which the response rate was best, but it was only about 38 percent. The overall response rate for the five-year follow-up mailed questionnaire was about 33 percent.

For the 11-year follow-up the downward slide continued--but fortunately at a decelerating rate; the overall respondent rate for the 11-year mailed questionnaire was down to 23 percent.

D. Special samples of nonrespondents

In order to maintain the probability-sample character of the data, it was necessary to get the nonrespondents represented somehow. This was accomplished by picking a probability sample of the nonrespondents and then contacting them directly, usually by phone, and interviewing them to get their answers to the questionnaire items. The weights for these cases had to be increased by an appropriate factor, to take account of the fact that each of these contacted nonrespondents to the mailed questionnaire represented not only themselves but also a large number of other nonrespondents, who weren't in the special sample.

The manner of selection of the special sample was quite similar in most of the follow-ups although in a few of them something slightly different (and usually a little more complex) was done. The most prevalent procedure was to arrange the nonrespondents in student number order and then to pick a "systematic" sample of them--i.e., every umpty-umpth nonrespondent--to get a sample of the desired size (usually about 2500 to 3000 cases). Modifications of this procedure occurred in the case of the 1-year follow-up of the grade 9 cohort, the 5-year follow-up of the grade 11 cohort, and the 11-year follow-up of the grade 9 cohort. In all three of these follow-ups, the procedure was exactly like the general procedure outlined above, with the exception that the following modifications were used.

For the 1-year follow-up of grade 9

The nonrespondents were stratified on the basis of whether they had been included in the 1963 retest sample or not. For the ones that had been, a higher sampling ratio was used.

For the 5-year follow-up of grade 11

The nonrespondents were stratified on the basis of whether they had responded to the one-year follow-up questionnaire. For the ones that had, a higher sampling ratio was used.

For the 11-year follow-up of grade 9

> The nonrespondents were stratified jointly on their "general academic aptitude composite" score and on the percentage of minority students who were enrolled in their high school in 1960. The sample was selected with selection ratios four times as high for high-minority-enrollment schools and twice as high for medium-minority-enrollment schools as for other schools, i.e., those with low minority enrollment and those that provided no information about minority enrollment. With those exceptions, the 9th-grade 11-year follow-up special sample was "self-weighted". (See Chapter 15, section C, for a brief explanation of "self-weighting".)

Obviously the special procedure that was used in the 5-year follow-up of grade 11 was designed to increase the proportion of the sample members on whom both one-year and five-year follow-up data would be available, so that results could be analyzed longitudinally, across follow-ups. Because the desirability of having follow-up data obtained at different times available on the same individuals is obvious, some readers may wonder why we didn't make more extensive use of the oversampling procedure--or use it to a more extreme degree. For instance why didn't we decide that to represent those 11-year nonrespondents to the mailed questionnaire who had also been nonrespondents to the 5-year mailed questionnaire, we would select as members of the 11-year special sample only those 5-year mailed-questionnaire nonrespondents who had been in the 5-year special sample? The logic seems impeccable; we would merely be taking advantage of the fact that the 5-year special sample already represented the 5-year nonrespondents not in the special sample. And the advantage is clear; our 11-year special sample would consist wholly (or almost wholly) of sample members on whom 5-year follow-up data were available. The answer to the question is that we did consider the possibility of selecting the 11-year special sample in the manner outlined above; we considered it very seriously. But we decided that before making a definite decision we should do one empirical check. We knew that at the time of selection the 5-year special sample was representative of all the nonrespondents to the 5-year mailed questionnaire. But were they *still* representative after the special follow-up interview experience? It seemed likely that they *would* be; after all it didn't seem likely that one interview--perhaps a 15-minute conversation on the phone--would change behavior six years later. But being empiricists we decided to make certain. There was just one kind of behavior on which we had data (the early returns to the 11-year follow-up mailed questionnaires) that could be compared for 5-year special sample members and 5-year nonrespondents *not* in the special sample. That behavior was the act of completing and returning the 11-year mailed questionnaire. Accordingly we computed a chi-square coefficient on the following four-fold table, based on all *nonrespondents* to the 5-year follow-up mailed questionnaire:

In 5-year special sample?	Respondent to 11-year mailed questionnaire?	
	Yes	No
Yes		
No		

Much to our surprise (or, at least, much to *my* surprise) the chi-square coefficient was statistically significant. A larger proportion of the 5-year special sample members than nonmembers returned the 11-year questionnaires. Since that one type of behavior had apparently been changed by the experience of the special-sample interview, we had to assume that perhaps other kinds of behavior subsequent to the interview had also been changed. We decided, therefore, not to apply the otherwise attractive idea of limiting the 11-year special sample to cases with 5-year follow-up data (plus those few 5-year special-sample members on whom interviews had *not* been completed).

As for methods of contacting members of the special sample, there has been continual experimentation. Among the different methods tried, either alone or in combination with others, in various follow-ups, have been the following:

1. Follow-up by the Regional Coordinator, by telephone interview.

2. Follow-up by the Regional Coordinator in a face-to-face interview. This approach was used only in the very early years of the project, and then only rarely. Whether the contact should be by phone or in person was left to the judgment of the Regional Coordinator, who generally would telephone except if the nonrespondent couldn't be reached by phone or if there was some other problem that couldn't be handled any other way than by a face-to-face interview.

3. Follow-up by Retail Credit*. Since this method was expensive it was generally reserved for nonrespondents on whom the Regional Coordinators--or sometimes the TALENT staff member who was doing the telephoning (as discussed in paragraph 4 below)--had to give up. But in the one-year follow-up of the 9th grade cohort, the Project Director decided to use Retail Credit's services for the entire special sample. It didn't work well (only two-thirds of the nonrespondents were interviewed) and was never tried again. (The decision to abandon this approach was good one, in the opinion of the author, for more reasons than just the low response rate. It seemed reasonable to suppose that our carefully selected Regional Coordinators, or other interviewers selected and trained by us, would be more skilled at the sort of interview required than would Retail Credit's employees, presumably trained for an entirely different type of telephone conversation.)

4. Phone calls from the Project TALENT office by junior staff members (research assistants, administrative assistants, coders), who were trained by senior staff members for this interviewing. This procedure of using staff members to make the calls was reserved, in the early follow-ups, for cases that the Regional Coordinators had given up on, or even sometimes, for cases that Retail Credit (see paragraph 3 above) had not been able to contact. In recent years, however, TALENT staff members have been used in preference to Regional Coordinators, except in areas where the Regional Coordinator has proven very effective in previous follow-ups.

---

* Retail Credit has recently changed its name to "Equifax".

5. Mailing of another questionnaire to a member of the special sample, after he or she has been contacted by phone and has agreed to fill it out promptly and return it. (This method is used only under special circumstances.)

For further information on methods of contacting members of the special sample see "Methodology of the 11-Year Follow-Up" (Rossi, Wise, Williams, and Carrel, 1976).

The completion rates for interviewing of the special sample, like those for the mailed questionnaire, have varied from year to year. The all-time high, 99.5%, was of course for the very first follow-up--the one-year follow-up of grade 12--when our address records were freshest and therefore mostly likely to be correct. Honors for the all-time low are shared by the five-year follow-up of grade 9 (under 63.0%) and the 11-year follow-up of grade 12 (just a bit over 63.0%).

E. Efforts to maintain and improve response rates

1. For mailed questionnaire

Completion rates depended, of course, not only on the state of the address file but also on the nature of the contact procedures used--as well as various factors out of our control. Such factors included the collective mood of the particular cohort of young people, and what proportion of them were unavailable because they happened to be in Viet Nam at the time of the follow-up.

Nevertheless there was quite a bit we could do to improve response rates. Some experimentation has been done to find out the optimum mailing schedule when there are several waves. For instance should there be two weeks between waves? Or a month? When should the reminder postcard go out? We have not used precisely the same mailing schedule in every follow-up of every cohort, but the schedule described in section A above represents our current thinking. It is the schedule that was used in the

11-year follow-up of grade 10. (A somewhat different schedule was used for grade 9, because of the operation of Murphy's well-known Law*.)

We have learned a lot about effective use of the mails, since our first follow-up. In the more recent follow-ups we have tried to use a reasonably sophisticated mix of first-class mailing (for [hopefully] speed and dependability) and bulk rate for economy.

For the 11-year follow-up of the grade 10 cohort, the mailing schedule was as shown in Table 18-1.

The Project TALENT News (the four-page "newspaper" that we mail to all members of the TALENT sample annually) is intended to maintain continuing interest in the project, and to help promote good will towards it. It is also an effective device for maintaining and updating address files. (Our current mailing procedure is to use bulk rate for mailing the newspapers, together with "address correction requested".)

2. For special sample

Efforts to improve the completion rate for the special-sample interviews have two distinct aspects: (a) developing better ways of locating hard-to-find people and (b) developing ways of securing the cooperation of the special-sample member, once he or she has been located. Getting their cooperation probably depends at least as much on the attitude and manner of the interviewer as it does on anything else. As for the procedures that have been developed for locating sample members, they are described in considerable detail in "Methodology of the Project TALENT 11-Year Follow-Up Study" (Rossi, Wise, Williams, and Carrel, 1976, Appendices B through D).

---

* Whatever can go wrong will.

TABLE 18-1. Mailing schedule and procedure for 11-year follow-up of Grade 10 cohort, and TALENT News

Item mailed	Wave	Mailing date	Interval in weeks	To whom sent	Method of mailing
TALENT News	-	6/18/73		Entire TALENT sample	Bulk mail Address correction requested
			13½		
Questionnaire	1	9/20/73		Entire 10th-grade cohort*	First class
			1		
Reminder postcard	-	9/27/73		Entire 10th-grade cohort	Bulk mail Address correction requested
			2		
Questionnaire	2	10/11/73		Entire 10th-grade cohort	Bulk mail Address correction requested
			7		
Questionnaire	3	11/29/73		Members of 10th-grade cohort who hadn't yet responded**	Bulk mail Address correction requested
			10		
Questionnaire	4	2/7/74		Members of 10th-grade cohort who hadn't yet responded (except that those for whom two dead-end address unknown returns were received were eliminated from this mailing)	Bulk mail Address correction requested

---

* Prior to mailing, the mailer tape was updated on the basis of name and/or address changes received as a result of the Project TALENT News.

** Prior to mailing, the mailer tape was updated on the basis of name and/or address changes received from prior mailings.

F. Content of special-sample questionnaire

Should the special sample questionnaire be the same as the mailed questionnaire or should it differ in some way? Four different levels of similarity may be recognized:

Level 1. The special-sample questionnaire uses an entirely different set of questions from that used in the regular questionnaire. (They are worded differently, and may have different options.)

Level 2. The special-sample questionnaire consists of a subset of the regular questionnaire's items, with minor changes in wording.

Level 3. The special-sample questionnaire consists of a subset of the regular questionnaire's items, identically worded.

Level 4. The special-sample questionnaire is identical with the regular questionnaire.

Gradually, we have progressed from Level 1 (the initial procedure) to Level 4 (the current procedure). Someone decided to use the Level 1 procedure in the one-year follow-up of grade 12. It worked out badly, for obvious reasons; if the special sample is supposed to provide information as to how the nonrespondents to the regular questionnaire would have responded to that questionnaire if they had responded, it is clear that the same questions have to be asked.

The idea behind Levels 2 and 3 (a subset of the regular questionnaire's questions) is that the amount of time that an interviewee can be expected to spend answering questions is limited. But the Level 4 approach allows us to have our cake and eat it too. It includes all the questions but they are asked in descending order of importance, and with the "sensitive" items of limited importance last, so that if the interviewee terminates the interview before answering all the questions, he has already answered all the questions we would have if the questionnaire had been abridged to start with. And if, on the other hand, the interviewee is willing to answer all the questions, he is given the opportunity to do so.

G. Ten becomes eleven; twenty equals seventeen?

It will be recalled from Chapter 1 that because of delays in funding, the 10-year follow-up, it became the 11-year follow-up instead. It now appears that the long-planned 20-year follow-up is about to turn into a 17-year follow-up, at the suggestion of NIE personnel.

Chapter 19. TALENT POST-DOCTORAL FELLOWSHIPS

During the academic year 1966-67, Project TALENT had a grant from the Office of Education to select three post-doctoral Fellows and to operate a post-doctoral training program for them, in computer and multivariate applications to educational research.

The three winners, selected from many applicants, were Dr. William A. Love (a social psychologist), Dr. David Kapel (an education major whose doctorate was in curriculum development), and Dr. Douglas Stewart (a sociologist). They received their doctorates from, respectively, the University of Texas, Temple University, and the University of Pittsburgh.

The entire senior staff of the project* participated in setting up the training program and providing the training, but the person who played the leading role in this activity was Paul Lohnes. Paul is a natural teacher, with a great enthusiasm for teaching and a real gift for conveying enthusiasm about whatever he is teaching--whether it happens to be multiple discriminant analysis, or Project TALENT procedures, or canonical correlations, or the elements of computer programming--to whoever his students happen to be. Whether the post-doctoral program was a success probably depends in large part on whether the three Fellows think it was a year well spent. To the best of our knowledge they did. Looking back retrospectively they were uniformly enthusiastic, regretting only that the program didn't last longer.

Their principal responsibility during the year was to complete a research study on a suitable topic, and write a report. Dave Kapel did a study that resulted in a report entitled "Effects of Negro Density on Student Variables and the Post-High-School Adjustment of Male Negroes"; it was an interesting effort, though unfortunately the usefulness of the report was severely reduced by the fact that data for the study were available for so few cases. (A smaller percentage of blacks than whites responded to the follow-up questionnaires.)

---

* The senior staff included: Bill Cooley, Paul Lohnes, Lyle Schoenfeldt, Bary Wingersky, Charles Hall, and Marion Shaycoft

Bill Love and Doug Stewart were so infected by Paul Lohnes' enthusiasm for multivariate analysis that they worked together to try to develop a method of determining what proportion of variance is accounted for by each canonical correlation, and to apply their method to the TALENT sociometric data. Their report is entitled "Interpreting Canonical Correlations: Theory and Practice".

After their year in Pittsburgh, at AIR, they scattered, of course. Doug Stewart traveled the least distance. He returned to the Sociology Department at his alma mater, the University of Pittsburgh. Dave Kapel joined the faculty of Glasboro State College, in New Jersey. (Since then he too has moved back to his alma mater, Temple University.) Bill Love accepted an appointment at Nova University, in Florida.

Chapter 20. LATER STUDIES

In this chapter we discuss studies by the TALENT staff that were done subsequent to those reported in Chapter 10. Though there is certainly no sharp distinction between the two groups of studies, possibly those discussed in the present chapter have less of a "necessary preliminaries" flavor and more of a "let's-get-on-with-it" flavor. But that distinction, if it exists at all, exists only tenuously. In any event, to really know what the various reports are like, the reader will have to look at the reports themselves. Therefore my advice is not to worry about the distinction between what went into Chapter 10 and what goes into Chapter 20, but instead, "let's get on with it"! As in Chapter 10, only major reports are discussed in this chapter. Furthermore it is limited to "mainstream Project TALENT studies". The term "mainstream Project TALENT" is used to designate that part of Project TALENT which was supported by OE or NIE.

A. The one-year follow-up studies

"The American High School Student"* (Flanagan, Davis, Dailey, Shaycoft, Orr, Goldberg, and Neyman, 1964) includes a presentation of preliminary results from the one-year follow-up of the grade 12 cohort. But until "Project TALENT One-Year Follow-Up Studies" (Flanagan and Cooley, 1966, 365 pages) was written there was nothing pulling together follow-up data from all four cohorts.

B. The MAP studies

Paul Lohnes factor-analyzed 62 variables--grade, sex, and 60 score variables (mostly cognitive)--in one analysis and 40 variables--grade, sex, and 38 noncognitive variables--in another. He came up with 13 factors in each analysis--grade, sex, and 11 "abilities factors" in the first, and grade, sex, and 11 "motives factors" in the other. He called these 22 factors the MAP factors (MAP = "Measuring Adolescent Personality"). MAP is a clever acronym; but as Truman Lee Kelley wrote many years ago, no two factor analysts agree on how factor analysis should be done--and therefore it is perhaps a foregone conclusion that the author of this History greatly prefers the factor analysis methodology she used in "The High School Years: Growth in Cognitive Skills" (see paragraph C below, or Chapter 11) and thinks it produced more meaningful results. (See also paragraph E below for yet another factor analysis of the TALENT battery.) Anyhow, two reports resulted from the MAP factor analysis--"Measuring Adolescent Personality" (Lohnes, 1966, 225 pages), and "Predicting Development of Young Adults" (Cooley and Lohnes, 1968, 233 pages). The first of these describes the factor analysis and the second applies the results to five-year follow-up data in an effort to relate factor scores to such things as career plans and career plan changes. The factor scores in many of the analyses are analyzed by multiple discriminant analysis, so that essentially the procedure used is a multivariate analysis of a multivariate analysis.

---

* See Chapter 10

It might be mentioned in passing, to forestall a deluge of requests by potential Data Bank clients (see Chapter 22) who wish to obtain and use MAP scores in their own research, that this is impossible. Paul Lohnes unfortunately did not present in the report the matrix of coefficients that defines the factor scores; nor did he present the complete tables of factor loadings* which would have permitted reconstruction of the matrix of coefficients; nor the original correlation matrix he used, which would have permitted reconstruction of the whole thing; nor did he provide the TALENT staff with any of these unpublished items. And since he did his own programming and managed his own computer operations personally, the MAP factors are irretrievably lost.

In view of my earlier comment that I prefer the factor analysis that I did on the 1963 retest, perhaps it should be pointed out that factor scores from that factor analysis aren't available either, but for an entirely different reason. They are not available because a matrix of coefficients to define factor scores was never computed. And the reason a matrix was not computed is that this author is opposed to such a use of factor scores. She thinks it just obscures results that would be clear and meaningful if presented in terms of the original variables. What value factor analysis has in substantive research lies in its use to gain a better understanding of the original variables--rather than in replacing those original variables with artificial abstractions called factor scores, that often don't even represent the extracted factors very well. Thus, in the opinion of the author, there is hardly ever any value in computing factor scores, even when the factor analysis from which they would be derived is sound and useful in helping to understand the underlying nature of the original variables.

---

* Only loadings of at least .35 were published.

C. The 1963 retest study

As described in Chapter 11, about 7500 students who had been tested in 1960 as 9th-graders were retested as 12th-graders in 1963. "The High School Years: Growth in Cognitive Skills" (Shaycoft, 1967, 376 pages) is a comprehensive report on the results and the methodology.

D. The five-year follow-up studies

The two major reports which came out of the five-year follow-up other than the already mentioned report by Cooley and Lohnes (1968) were "Five Years After High School" (Flanagan, Shaycoft, Richards, and Claudy, 1971, 492 pages plus a separate 359-page Appendix) and "The Career Data Book" (Flanagan, Tiedeman, Willis, and McLaughlin, 1973, 397 pages).

"Five Years After High School" not only presents substantive results from the follow-up but also describes some methodological innovations (e.g., propinquity analysis, an idea developed by Shaycoft).

"The Career Data Book" is discussed in paragraph F below.

E. Ted Cureton's two TALENT studies

Edward E. (Ted) Cureton has never been on the Project TALENT staff, but he has been a consultant and has been sufficiently interested in the project that he arranged to do two important studies using the data. One was an equating study and the other was a factor analysis.

The equating study, which was carried out on Knoxville students who had already taken the TALENT battery, involved administering three other batteries to these students, so that tests in those batteries could be equated to the TALENT tests. The three batteries were the FACT Battery, the Differential Aptitude Tests, and the General Aptitude Test Battery. The methodology and results are reported in "Project TALENT Tests as Anchors for Equating Other Tests" (Cureton, E. E., 1973, 60 pages).

The factor analysis was based on the same data as the equating, plus one additional battery, the Essential High School Content Battery. The methology and results are reported in "A Factor Analysis of Project TALENT Tests and Four Other Test Batteries" (Cureton, E. E., 1968, 79 pages).

F. The Career Data Book and the supplement to it

"The Career Data Book" (Flanagan, Tiedeman, Willis, and McLaughlin, 1973, 397 pages) presents TALENT score profiles for groups of people in various career fields on the basis of the five-year follow-up. These are intended for use in career guidance.

The Career Data Book supplement, which is entitled "Using the TALENT Profiles in Counseling" (Rossi, Bartlett, Campbell, Wise, and

McLaughlin, 1975, 216 pages), does essentially the same thing, except that 11-year follow-up data (for the Grade 11 and Grade 12 cohorts) are used instead of 5-year data, the career groups are defined a little differently, and the profiles are presented in a slightly different format.

G. The 11-year follow-up studies

In addition to the Career Data Book supplement described above, the three chief substantive reports that have come out of the 11-year follow-up thus far are "The American Citizen: 11 Years After High School" (Wilson and Wise, 1975, 163 pages), "Post High School Education and Career Development" (Yen and McLaughlin, 1974, 90 pages), and "Careers in Science: A Project TALENT Study" (Shaycoft, 1975, 132 pages).

The Yen-McLaughlin report compares students with varying amounts and kinds of post-high-school education, with respect to such career-related variables as salary, job satisfaction, and employment rate. (Socioeconomic status and academic aptitude are used as control variables.)

"The American Citizen" presents response distributions for the 11-year follow-up questionnaire items, for the Grade 11 and Grade 12 cohorts.

The "Careers in Science" report, as its title suggests, is concerned with the use of Project TALENT data in choosing a career in science, and with problems of career guidance in high school, for potential scientists. Although funds for this study were provided by NSF rather than by OE or NIE, it is still, in a sense, a part of "mainstream Project TALENT", since the NSF funds were administered by OE (and then later by NIE) via an interagency agreement. (See Chapter 17.)

In the next chapter some further, quite varied, uses of Project TALENT data are discussed.

Chapter 21. SOME OTHER USES OF TALENT DATA

Over the years a great many studies involving Project TALENT data in one way or another have been carried out (and a great many that might have been carried out haven't been). Researchers from numerous disciplines--not only educational researchers but also psychologists, economists, sociologists, home economists, library science specialists--have used TALENT data in their studies.

This chapter is a potpourri of studies of various kinds which have very little in common except that they all involve Project TALENT. The chief purpose of this chapter is to give some idea of the breadth of applications. Again, as in Chapters 10 and 20 we shall make no effort to report the findings of the various studies.

In the sections that follow, the studies are classified into loose categories (since they don't fall into tight categories very well).

A. Equating studies

Three of the five equating studies have been briefly discussed in previous chapters. The following instruments have all been equated to appropriate TALENT Tests or composites in order to obtain norms approximately equivalent to those that would have been obtained had the test been administered to the TALENT sample.

1. For the Air Force:

Armed Forces Qualification Test (AFQT): 1 variable

Air Force Officer Qualification Test (AFOQT): 27 variables

Airman Qualifying Exam (AQE): 4 variables

2. For the Navy:

Navy Basic Test Battery: 5 variables

3. Knoxville equating

Equating of tests in three other batteries (FACT, DAT, GATB) to TALENT tests. (This equating was done by Ted Cureton under a grant from OE.)

4. Inter-American Tests

This battery by Herschel T. Manuel was equated to the TALENT tests.

5. PCG equating

Tests of the Planning Career Goals (PCG) battery have been equated to the TALENT tests. For a discussion of the PCG battery, see Chapter 23.

B. Knoxville follow-up studies

Louise Cureton has done three separate studies involving special data collection efforts for the Knoxville sample. The first was a study of the correlates of juvenile delinquency; the second was a study of welfare dependency; and the third, a study of alcoholic problems among adolescents in the TALENT sample, and their parents.

Other researchers besides Louise Cureton have also done studies for which they found the Knoxville sample peculiarly useful. For instance John Claudy, David Gross, and Rebecca Strause (three AIR staff members) prepared a report entitled "A Study of Married Couples in Knox County, Tennessee" (1974).

C. Other studies of marriage, the family, and parenthood

Claudy, Gross, and Strause also did another population study (this one not limited to the Knox County population) which resulted in a report entitled "Family Size, Birth Order, and Characteristics of Young Adults" (1974).

Josefina J. (JJ) Card* has completed a study (Card, 1977) of the consequences of adolescent childbearing for the young parent's future personal and professional life. She is now doing a related study, on long-term consequences for children of adolescent parents.

D. Changes in norms

John Flanagan has personally carried out two studies which involved going back to a sample of the TALENT schools 10 or more years later, giving one or more of the same tests, and comparing performance with that in 1960. The findings of both studies have had considerable bearing on the current controversy concerning test score decline.

The first was the "Progress in Education" study. This involved, among other things, going back to 134 of the TALENT schools in 1970, giving the Reading Comprehension Test again (along with some of the Student Information Blank items) to all 11th-grade students or a representative sample of them, and comparing the results with those achieved in 1960. The resulting report is entitled "Progress in Education: 1960-1970" (Flanagan and Jung, 1971). This study was funded in part by NIH through general research support funds.

In 1975, in connection with the equating of the PCG battery (see Chapter 23) to the TALENT battery it was necessary to administer both batteries to about 28 schools, and for the purpose some schools from the original TALENT sample were used. Findings relevant to the test-score-decline controversy are reported by Flanagan in "Changes in School Level of Achievement: Project TALENT Ten and Fifteen Year Retests" (Flanagan, 1976). This study was financed by AIR itself, rather than by a government grant or contract.

E. Quality-of-life studies

In recent years, John Flanagan has personally directed two quality-of-life studies that utilized TALENT cases and TALENT data (as well as new follow-up data). The first of these studies, which was done by John Flanagan and Sandy Wilson, was based on about 250 TALENT cases who had Army experience; it was sponsored by the Army Research Institute (ARI). Results are reported in "Quality of Life as Perceived by 30-year-old Army Veterans" (Wilson, Flanagan, and Uhlaner, 1975). The second was a study of the quality of life of 30-year-olds; it was based on a 1,000-case subsample from TALENT's probability sample of 15-year-olds (500 men and 500 women), who were interviewed 15 years later. The principal purpose of the study was to help formulate educational goals. It was sponsored by NIE. The results are reported in "An Empirical Study to Aid in Formulating Educational Goals" (Flanagan and Russ-Eft, 1975).

---

* She is a member of the AIR staff.

F. Follow-back studies

Several of the Data Bank studies have been of the follow-back type. Unfortunately, however, these studies have been neither as numerous nor as successful as the potential of the method might lead one to expect. The reasons have been discussed in Chapter 13.

G. A guidance study for NSF

This study, unlike the Shaycoft study (1975) mentioned in section G of Chapter 20, was done directly for NSF rather than being administered by OE or NIE. But like the Shaycoft study, it is concerned with career guidance in secondary schools. The report is entitled "Development of Scientific Careers: The High School Years" (Gilmartin, McLaughlin, Wise, and Rossi, 1976, 216 pages). Even though this study and the Shaycoft study deal in large part with the same problem, their approaches and methodologies, and the specific questions they seek to answer, are entirely different.

H. Sex differences in achievement

J. J. Card (see section C above) is currently doing a study of sex differences in realization of achievement potential. It is sponsored by NIMH.

I. All-volunteer military forces

In a study done by John Claudy, Project TALENT data were analyzed to provide insight into problems associated with the establishment of an all-volunteer force.

J. Project PLAN

Project PLAN was a major enterprise conceived and directed by John Flanagan. PLAN is an acronym for "Program for Learning in Accordance with Needs". The purpose of the project was to develop, evaluate, and demonstrate a comprehensive system of individualized education for use in grades 1 through 12. One important goal of the program is to help the students develop and clarify their educational and career plans; educational programs are then individualized partly on the basis of these plans. This aspect of the program leans very heavily on TALENT data.

The development of the PLAN materials was financed primarily by Westinghouse Learning Corporation, but also to a substantial extent by AIR. The 14 participating school systems also made a substantial investment, in the form of contributed staff time.

K. Data Bank studies

The variety and scope of these studies is considerable. The reader is referred to Chapter 22.

L. Miscellany

Over the years, Project TALENT staff members have presented a total of over 100 TALENT-related papers at professional meetings, and have had about 60 articles published, all of which have contributed to the breadth of scope of the project.

M. Undone studies

In the first paragraph of this chapter there was a reference to studies "that might have been carried out [but] haven't been". In section F above, it was indicated that some potential follow-back studies might have fallen in this category.

Among other kinds of uses of TALENT data that are potentially promising but have hardly been tried are studies involving the development of special score scales, tailor-made for a particular investigation. Such studies could be done using the 4%-sample tape, which, it will be recalled from Chapter 12, contains item response data for all items. Unfortunately, however, the 4% tape has not yet been put in as good shape as it should be, nor in as good shape as would be necessary for many studies of the type indicated. For a discussion of the 4%-sample tape and related matters, see Chapter 13.

For many potential studies requiring item-response data, the 1963 retest tape would theoretically be a good substitute for the 4%-sample tape, if the researcher were willing to limit his study to grade 12 cases. Unfortunately, though, the retest tape is in no better condition right now than the 4% tape. (See Chapter 13 again!)

We still have hopes that it will be feasible to get both tapes in first-rate shape. It would be helpful not only to potential Data Bank clients but to Project TALENT staff as well. (Actually, it would be helpful to the entire educational research community to have these resources available in first-rate condition, including first-rate documentation.)

Chapter 22. THE DATA BANK

The term "Data Bank study" loosely means, in TALENT terminology, a study carried out by a researcher who is probably (though not necessarily*) not a member of the Project TALENT staff and who uses TALENT data (printouts or tapes) especially prepared to his specifications, for use in his own research.

It is true that some project staff members, particularly in recent years, have been treating the term as if it had a much broader meaning, subsuming within itself all TALENT research, and all uses of any of the TALENT tapes. But the term is not used that way anywhere in this History.

A. History of the Data Bank

Although the data bank concept was embedded in the early plans for the project, the Data Bank operation did not get organized immediately. In the first place nothing could be done along those lines until there were tapes ready to use in data analysis.

However by the middle of 1962 things were in good enough shape that we could accept our first data bank contract. This was a contract to provide J. Alan Thomas, then at Stanford, with data for use in an analysis of school and community characteristics as predictors of test scores and other variables. Thomas provided the input data on community characteristics; Project TALENT provided all the other data. There were about four other data bank studies done during the Pittsburgh phase. One was for the Institute of Life Insurance. A second was for the National Merit Scholarship Corporation, where John Holland was doing a study of National Merit Scholarship finalists. The third was carried out for the Russell Sage Foundation, where David Goslin was doing a study of the impact of testing. Two large reports resulted: a technical report (Brim, Goslin, Glass, and Goldberg, 1964) and a book reporting the substantive results (Goslin, 1963).

---

* Even John Flanagan carried out a Data Bank study (in connection with Project PLAN)--proof positive that TALENT staff is not necessarily excluded from the ranks of Data Bank users!

The fourth data bank study was done for NSF, and it was on academic degree projections. Actually this fourth study was not really a "data bank study" in terms of the definition, since it was not just a matter of obtaining computer output and presenting it to the sponsor. A member of the Project TALENT staff, specifically Izzy Goldberg, together with John Dailey, helped develop the research plan, analyzed the data, and wrote the report.

But though even with the NSF project excluded at least four studies of the data-bank type were completed during the Washington phase, the "Data Bank" (with a capital D and capital B) was not formally organized until the move to Pittsburgh. As mentioned in Chapter 8, Dick Holdeman became the Data Bank Coordinator when he joined AIR in the fall of 1967. Since his departure a year later there have been three Data Bank Directors: Lyle Schoenfeldt, John Claudy, and Don McLaughlin.

B. Extent of use

About 140 Data Bank studies have been completed, since the Data Bank began its formal existence, in the fall of 1964. Of these about 40% have dealt with education, 10% with career development in general, 5% specifically with science careers, 9% with family and intergenerational effects, 7% with methodology, and the remaining 29% with a wide variety of miscellaneous topics, including, for instance:

Participation in athletics
Gifted students
Black male adolescents
Teachers
Airline stewardesses
Cognitive style
Mental patients
Suicide
Social mobility.

Many of the Data Bank clients are repeaters. For instance J. William Asher, who is at Purdue, and is possibly the heaviest Data Bank user, has been responsible for at least half a dozen studies based on TALENT data; the studies have been mostly concerned with driver education and auto accidents.

John Folger, Alan Bayer*, and Helen Astin, who were on the staff of the now defunct Commission on Human Resources and Advanced Education, also did many Data Bank studies, which dealt mostly with educational attainment beyond the high school level, and the factors to which it is related. As a matter of fact in terms of institutions rather than individuals the Commission on Human Resources was probably the heaviest user of the Data Bank--except for AIR itself and the University of Pittsburgh--having commissioned at least nine separate projects. (If it seems to the reader that the counts of projects are expressed in slightly vague terms--e.g. "'at least' nine", etc.--that is because in the case of closely related projects it is not always clear where one leaves off and another one starts.) The three former Commission on Human Resources staff members--Folger, Astin, and Bayer--produced a book entitled "Human Resources and Higher Education" (1970), which was based in part on TALENT data. After Helen Astin left the Commission on Human Resources, she continued to use the Data Bank, too, in her new role as a member of the staff of the Bureau of Social Science Research.

Another multiple user is John Hause, of the University of Minnesota; he is among the economists who view TALENT data as a useful resource for their research, and his interest in the project has continued over a very long period.

Two former directors of Project TALENT--John Dailey and Bill Cooley--were users of the Data Bank's services after departing from AIR.

---

* Al Bayer subsequently joined the AIR staff (see Chapter 8, section A8).

Christopher Jencks used TALENT data in his well-known study leading to his book entitled "Inequality" (Jencks et al., 1972). He also used the outcomes of a Data Bank study in an article of major importance entitled "Effects of High Schools on Their Students" (Jencks and Brown, 1975).

C. Charging for Data Bank services

Data Bank projects have been charged for at cost right since the beginning. As a matter of fact except under special circumstances there is no charge for preliminary consultations that take place with a researcher who decides not to go ahead with a Data Bank study. Thus, since the fixed costs of operating the Data Bank have to be covered, charges to the users have to be set high enough to cover the costs of possibly lengthy consultations that do not lead to Data Bank projects.

Before a project is actually started a cost estimate is prepared and a contract is signed. Most of the contracts call for charges on a cost-reimbursement basis, but (usually at the client's request) some contracts have been written on a fixed-price basis.

D. Policy on use of TALENT Battery

Some potential Data Bank clients wish to do research that requires administration of part (or all) of the TALENT battery to a new group. It will be recalled from Chapter 3, section C, that it had been agreed that a limited number of copies of the tests would be preserved for use in possible future research, but that the battery would not be used in competition with commercial tests. In line with this understanding, we have permitted administration of TALENT tests or inventories if the proposed use falls in one of the following categories:

1. The purpose is to equate some other test (commercial or otherwise) to the TALENT battery, so as to be able to take advantage of the TALENT norms or (indirectly) of the TALENT follow-up data.

or

2. The proposed research requires the use of one or more TALENT tests because no instrument available commercially would meet the needs of the research as effectively.

or

3. The proposed research fits in with our planned program for Project TALENT research, and will further the progress towards those goals.

It will be observed that routine operational use does not qualify under any of the three categories. Operational use is not permitted and never has been.

When permission is given to a researcher to use TALENT tests, either of the following opticns may be arranged:

1. The researcher may borrow an appropriate number of test booklets and accessories from Project TALENT, paying only a nominal "rental fee" for them. The fee is set to cover our costs, including the costs of storing the booklets, shipping costs, etc.
2. The reseacher may instead prefer to get permission to reprint the tests or other instruments he wishes to use. (In this case there is no rental fee.)

E. Policy on use of Data Bank services

Two major and continuing problems which have had to be dealt with in operating the Data Bank have been the necessity of maintaining the confidentiality of data, and the entirely separate problem of whether tapes containing raw data (after suitable purging of identifying information) should ever be released to Data Bank users so that they can do their own analyses or whether, instead, the Project TALENT staff (specifically the Data Bank staff) should be responsible for getting the computer analyses done, and should provide the Data Bank client only with data analysis printouts and/or data analysis summary output tapes.

These two problems are discussed in turn, below.

1. Maintaining confidentiality

Perhaps the one immutable principle in operation of the Data Bank has been that no data or tapes will be released in a form that violates the confidentiality of either the 1960 data or any follow-up data for any individual, or for any participating secondary school, or school system, or community, or state. Data cannot be released for any geographical unit smaller than region of the country, except where the geographical unit is neither identified nor identifiable. Furthermore we have honored our commitment not to release data that would permit a comparison of the three categories of school control: public, parochial, and other private.

One of the things these commitments mean is that before releasing tapes containing TALENT data for individual students, they must be purged of identifying (or partially identifying) information such as student's name, date of birth*, address, mother's maiden name, father's first name, and school attended. (Our regular school codes have to be replaced by dummy codes.) Furthermore we have to be careful that analyses (either printouts or output tape) are not in such a form that individual schools, communities, states, or other entities about which data should not be released cannot be identified even by inference.

2. Raw data tapes versus data analysis printouts

The initial policy of the Data Bank was that raw data tapes would not be released, even after removal of identifying information, except under very special circumstances. Data tapes would be released only if in the opinion of Project TALENT's special Data Bank Policy Review Committee (John Flanagan, Marion Shaycoft, the Data Bank Director, and the Project Director) it wouldn't endanger our commitments on confidentiality and if furthermore it was advisable for at least one of the following reasons:

---

* Age in years is retained on the tape.

1. Project TALENT didn't have and couldn't develop, or acquire economically, a computer program suitable for the type of analysis the researcher wished done.

2. The computer schedule or the schedule of Data Bank personnel was so crowded that it wouldn't be feasible to complete the analysis within a reasonable time.

We wished to keep tape release to a minimum because we felt that providing some users with tapes instead of doing their analyses for them would raise the costs of those Data Bank clients who didn't have convenient access to a computer and therefore needed to be provided with printouts, not tapes. Furthermore we felt that we could run analyses of TALENT data more economically than other people could because we were more familiar with the data and understood it better.* Although those reasons for preferring other alternatives to the release of data tapes still have some validity, our policy has loosened up considerably on this issue since about 1974. The reason for this is that many of the Data Bank users are graduate students who, wishing to base a master's thesis or a doctoral dissertation on TALENT data, don't have funds at their disposal to pay for extensive Data Bank analyses, but do have enough funds to pay for a data tape, after which they can get free time on their university's computer. This was one of the reasons a committee appointed by NIE to review various longitudinal studies, including Project TALENT, in the fall of 1973 recommended letting researchers have easier access to TALENT data tapes. We have changed our policy in accordance with this recommendation, although we recognize that there is a trade-off involved. In liberalizing the policy on data tape release we are making costs lower for some (not all) of those clients who receive tapes (mostly just for the ones that have access to free computer time at their own universities), but, since fixed costs of operating the Data Bank have to be covered, the costs become higher for the other clients.

---

* See section F of this chapter. Also see Chapter 13 again!

F. Handbooks, brochures, etc.

It was obvious when the Data Bank was formally initiated that some sort of brochure or handbook would be needed, outlining the kinds of data available, the services offered, and the administrative arrangements.

The first effort along these lines was a 28-page 8½" x 11" handbook entitled "The Project TALENT Data Bank" produced in the spring of 1965, when Dick Holdeman was in charge of the Data Bank. In 1968, when the Data Bank had its next director, Lyle Schoenfeldt, a small brochure was prepared that folded to fit in a standard business envelope. This was intended to supplement, not replace, the handbook prepared three years earlier.

A few years later, as use of the Data Bank grew, it became apparrent that the handbook would have to "grow", too. A larger more comprehensive one was needed. Consequently a new handbook was prepared. John Claudy, the Data Bank Director at the time, prepared much of the first draft; the general outline and the content were a joint effort of the senior staff (Claudy and Shaycoft writing; Shaycoft, Tiedeman, and Flanagan reviewing). John Claudy was responsible for the format. This handbook, 8½" x 11" again, was three-and-a-half times as long as the first one, even if one excludes the pages on which the follow-up questionnaires are reproduced,* leaving 96 pages.

---

* The follow-up questionnaire pages have to be excluded in comparing successive Data Bank handbooks, because the number of follow-up questionnaires that have been produced and have to be reproduced keeps increasing, and because the number of pages they fill depends on whether the reproduction is page-for-page (which usually produces a clear easy-to-use reproduction) or two-to-a-page (which is a slightly more economical mode but considerably less readable).

It will be recalled from Chapter 17 that the longitudinal study review panel that met on October 18, 1973 to review Project TALENT recommended comprehensive data tape documentation so that persons not on the TALENT staff who acquired TALENT data tapes, and other Data Bank clients who wished to plan their data analysis efficiently and economically [i.e., without having to make a trip to Palo Alto to confer with Project staff and without having to make numerous lengthy and therefore expensive long-distance calls to the Project TALENT office] could do so.

As mentioned in Chapter 17, the October 18 meeting and other considerations resulted in a proposal submitted to NIE that called for, among other things, the reformatting and improved documentation of the data tapes. An associated activity was the production of an updated Data Bank Handbook. The grant that was awarded to do this work was incorporated in a larger grant which started in July 1975. Marion Shaycoft drew up plans for a Data Bank Handbook that would meet the needs not only of Data Bank clients and potential clients but also of the programmers for the project and researchers on the project staff.

The responsibility was shifted to Laurie Wise and a complete draft of the revised handbook was submitted to NIE on July 16, 1976 as scheduled. After 6 to 8 weeks of review the draft was accepted without a single negative comment. The final printing was delayed until March 1977 in order that some internal suggestions could be incorporated.

Chapter 23. BRIDGING THE GAP:
The PCG (Planning Career Goals) Battery

It will be recalled that very early in Project TALENT--in fact while it was still in the planning stage--a commitment was made that the Project TALENT battery would be reserved for research purposes only, and would not ever be used operationally. This commitment was made for very good reasons, but it created problems nevertheless. Our five-year and 11-year follow-ups had produced data that would make the TALENT battery, or a battery more or less parallel to it, extremely useful in career guidance of high school students. But the schools couldn't use it very well because there was no test battery available that they could use operationally. We became thoroughly convinced of the difficulty of using other tests, more or less parallel to TALENT tests, but unequated to the TALENT battery, when efforts were made to use the Career Data Book (see Chapter 20, section F) under those circumstances. The Career Data Book had been widely distributed (every one of the over 1300 schools that had participated in Project TALENT had received one, as well as many counseling specialists on university faculties), and we had been inundated with reports that it was difficult to use them without having TALENT test scores for the students (or scores on directly parallel tests that could be expressed as TALENT equivalents), so that the Career Data Book data would apply directly. Because of the difficulties encountered in using the Career Data Book, it was arranged that Don McLaughlin and Wendy Yen would run a series of workshops for counselors on "How to Use the Career Data Book". But even these workshops, though carefully planned and expertly run, didn't help much. Career Data Book users still felt they were handicapped by the fact that scores for their students were not available on the TALENT battery. But the TALENT battery was barred from operational use by the commitment made many years before; and no parallel battery existed. It was almost a "Catch-22" situation--but not quite! A classic "Catch-22" situation has no possible solution; but the dilemma described above could be, and was, resolved.

A. Resolution of the dilemma

It was clear that the only sensible solution was to develop a battery, parallel to the TALENT battery, that could be used in the schools. Early in 1974 it was therefore agreed that AIR would undertake the construction of such a battery.

Since a two-day battery would be impractical from the schools' point of view, the battery would be designed to be administered in a single day. This meant that not everything in the TALENT battery could be administered. But that did not strike us as a serious problem. After all, after more than a dozen years of doing research on and with the TALENT battery, we felt we knew (and still feel that we know) which components of the battery were the most important and most useful ones. Those were the components we would concentrate on, in developing an abridged battery, to be given in a single day.

The plan called for equating the new tests to the corresponding tests of the TALENT Battery, by an equipercentile procedure, so that TALENT norms and TALENT follow-up data could be translated into norms and follow-up data for the new battery.

It was agreed that the new battery would be developed by the only three members of the then-current staff in Palo Alto who had been involved in development of the original battery; they were John Flanagan, Marion Shaycoft, and Mary Willis.

CTB/McGraw-Hill, in Monterey, California, was interested in publishing and distributing the new battery. In due course, arrangements were made with them.

That left us with hardly any problems except to plan the battery, develop it, and name it!

B. Planning the battery

Let's start with the hardest problem: naming the battery! It must have been the hardest problem; it certainly took the longest to solve. From the time that it was first decided to construct the battery to the time when the name was finally settled on, which was at almost the last minute, very shortly before materials had to go to press in final form, over a year had elapsed, and we had gone through about four separate and distinct battery names--adopted them for a couple of weeks or a couple of months, and then rejected them. One thing was clear. The name of the battery had to reflect the fact that it was intended primarily for use in career guidance. The name we finally settled on wasn't an acronym for anything (all the previous names had been) but it did make the purpose of the battery clear. The name was:

Planning Career Goals (PCG)

Fortunately we didn't have to postpone work on the other problems (the "easy" ones) while we struggled with the problem of battery name. We proceeded immediately with planning the battery.

It was decided that the battery would be scored primarily in terms of the 12 career groups into which John Flanagan had classified careers in connection with his research on Project TALENT data, and around which certain aspects of Project PLAN had been built. These 12 career groups had been used extensively in TALENT research, and had been found to work well. (The 12 groups are shown in Table 23-1.)

Accordingly it was decided that the Information Test to be developed and the new Interest Inventory would each yield 12 scores--one for each of the 12 career groups. In addition to the 12 interest scales and the 12 information scales, 10 ability measures would also be developed.

TABLE 23-1.

List of the Occupations Included in Each of the 12 Career Groups

1 ENGINEERING, PHYSICAL SCIENCES, MATHEMATICS, ARCHITECTURE

- Aeronautical Engineer
- Architect
- Chemical Engineer
- Chemist
- Civil, Hydraulic Engineer
- Computer Engineer
- Computer Systems Analyst
- Electrical, Electronic Engineer
- Mechanical, Automotive Engineer
- Mining, Petroleum, Metallurgical Engineer, Engineering Manager
- Physical Scientist: Geologist, Meteorologist
- Physicist
- Statistician, Mathematician
- Teacher: College Science, Mathematics

2 MEDICAL, BIOLOGICAL SCIENCES

- Conservation Specialist: Wildlife, Forestry, Agriculture
- Dentist
- Dietitian
- Life Scientist: Biologist, Microbiologist, Biochemist
- Medical, Dental Technologist
- Medical Specialist: Veterinarian, Optometrist, Osteopathic Physician
- Pharmacist
- Physical Therapist, Occupational Therapist
- Physician
- Registered Nurse

3 BUSINESS ADMINISTRATION

- Accountant, Auditor, Comptroller
- Advertising Worker
- Banking, Finance Worker
- Business Executive, Manager
- Certified Public Accountant
- City Planner
- Commercial Pilot
- Computer Programmer
- Efficiency Expert
- Hospital Administrator
- Manufacturing Manager
- Market Analyst
- Officer: U.S. Air Force, Army, Coast Guard, Marines, Navy
- Personnel Administrator
- Public Administrator
- Public Relations Worker
- Retail Store Buyer
- Sales Manager
- Stockbroker
- Wholesale, Retail Trade Manager

4 GENERAL TEACHING, SOCIAL SERVICE

- Clergy, Other Religious Workers
- Counselor
- Educational Researcher, Evaluator
- School Administrator
- Social Worker
- Speech Teacher, Speech Therapist, Reading Specialist
- Teacher:
  - Art
  - Elementary School
  - Of the Handicapped
  - High School Commercial Education
  - High School English
  - High School Foreign Language
  - High School Home Economics
  - High School Mathematics
  - High School Physical Education
  - High School Science
  - High School Social Studies
  - High School Trade, Industrial Education
  - High School (Other Subjects)
  - Music
  - Preschool
  - (Not Elsewhere Classified)
- Vocational Counselor

5 HUMANITIES; LAW; SOCIAL, BEHAVIORAL SCIENCES

- College Administrator
- Editor
- Journalist
- Lawyer
- Librarian
- Psychologist
- Social Scientist: Physical Anthropologist, Economist, Geographer, Historian, Political Scientist, Sociologist
- Teacher:
  - College English, Foreign Language
  - College Social Science
  - College (Other Subjects)
- Writer (not including Journalist)

6 FINE ARTS, PERFORMING ARTS

- Commercial Artist
- Interior Decorator
- Musical Artist
- Performing Artist

7 TECHNICAL JOBS

- Computer Operator
- Computer Services Repairer
- Dental Hygienist
- Drafter
- Electronics Technician
- Laboratory Technician
- Medical, Dental Technician
- Photographer
- Surveyor

8 PROPRIETORS, SALES WORKERS

- Automobile Sales Associate
- Business Supervisor
- Farm, Ranch Owner or Manager
- Independent Contractor
- Independent Proprietor
- Insurance Sales Agent
- Purchasing Agent
- Real Estate Sales Agent
- Route Delivery Sales Associate
- Salesclerk, Checker
- Sales Associate (Including Sales Engineer)

9 MECHANICS, INDUSTRIAL TRADES

- Airplane Mechanic
- Appliance Repairer
- Automobile Mechanic
- Clothing and Fashion Trades Worker
- Diesel, Other Specialized Mechanic
- Electrician
- Industrial Machine Repairer
- Machinist
- Office Small Machine Servicer
- Printing Trades Worker
- Telephone Installer, Repairer

10 CONSTRUCTION TRADES

- Bricklayer, Mason, Roofer, Plasterer
- Carpenter
- Explosives Handler, Construction Machinery Specialist
- Metal Trades Worker
- Miner, Oil Well Driller
- Plumber, Pipefitter
- Road Construction Worker
- Shift Leader

11 SECRETARIAL-CLERICAL, OFFICE WORKERS

- Accounts, Records Worker
- Bank Clerk, Teller
- Bookkeeper
- Cashier
- Clerk
- Keypunch Operator
- Legal Secretary
- Medical, Dental Secretary
- Office Machine Operator; Shipping, Stock, Mail Clerk
- Office Receptionist; Hotel, Reservation Clerk
- Secretary
- Secretary-Bookkeeper
- Stenographer, Court Reporter
- Telephone Operator
- Typist, Transcribing Machine Operator

12 GENERAL LABOR; PUBLIC, COMMUNITY SERVICE

- Barber
- Butcher, Meat Cutter
- Domestic Worker
- Enlisted Personnel: U.S. Air Force, Army, Coast Guard, Marines, and Navy
- Farm, Ranch Worker
- Firefighter
- General Laborer
- Hairdresser, Manicurist
- Mail Carrier
- Medical Aid
- Police Officer
- Practical Nurse
- Railroad Worker
- Service Worker: Waiter, Hospital Attendant, Usher
- Taxi, Bus, Truck Driver

Table 23-2 lists the PCG instruments, shows the number of items in each, and indicates the corresponding TALENT instrument. It will be noted that for most of the scales, except those few that are already very long in the TALENT battery, the number of items is somewhat greater in the PCG than in the corresponding TALENT test. This was planned in order to give the PCG scales the additional reliability they would need to make them useful in individual guidance, not just in research.

In general the reliability of the PCG tests turned out to be about equivalent, on a per-item basis, to that of the TALENT tests.

C. Developing the PCG Battery and equating it to TALENT

The three writers of the PCG battery split the test development work among themselves as follows (arranged in ascending order of volume:

Mary Willis--primarily responsible for:
- Reading Comprehension Test
- Creativity Test
- Mechanical Reasoning Test

Marion Shaycoft--primarily responsible for:
- Vocabulary Test
- English Test
- Visualization Test
- Abstract Reasoning Test
- Quantitative Reasoning Test
- Mathematics Test
- Computation Test

John Flanagan--primarily responsible for:
- Information measures
- Interest Inventory
- All manuals
  - Examiner's manual
  - Student's Guide (Planning Your Career)
  - Career Handbook
  - Counselor's Handbook
  - Technical Bulletin #1
  - Planning Profile
  - Planning Report
- Arrangements for and execution of equating
- Other arrangements

TABLE 23-2. PCG and Corresponding TALENT Instruments

P.C.G. Scales		Corresponding TALENT Scales	
Name of instrument	No. of items*	No. of items*	Name of instrument
Ability Measures			
1. Vocabulary	60	30	Vocabulary
2. English	48	113	English Total
3. Reading Comprehension	40	48	Reading Comprehension
4. Creativity	24	20	Creativity
5. Mechanical Reasoning	25	20	Mechanical Reasoning
6. Visualization	25	16	Visualization in 3 Dimensions
7. Abstract Reasoning	22	15	Abstract Reasoning
8. Quantitative Reasoning	22	16	Arithmetic Reasoning
9. Mathematics	24	24	Introductory H.S. Mathematics
10. Computation	76	72	Arithmetic Computation
Information Measures	(240)	395	Information Test
12 Career Groups	20 ea		
		252	Part I
		10-24	15 separate subscores
		143	Part II
		2-15	23 separate subscores
Interest Inventory	(300)	(205)	Interest Inventory
1. Occupations	(104)	(122)	I. Occupations
2. Occupational Activities	(99)	(83)	II. Activities
3. Current Activities	(97)		
12 Career Groups	25 ea		
		3-18	17 Occupational Categories

* Numbers in parentheses do not correspond to scales on which scores are obtained.

The tests were developed in draft form, tried out, item analyzed, revised and put in final form, and then equated to the corresponding TALENT tests. The equating took place in 1975; it was based on data obtained in 28 schools where both PCG and TALENT tests were administered. Of the 28 schools, 17 had been in the TALENT sample in 1960. It was from these 17 schools that the "15-year retest data" on which John Flanagan reported in his paper entitled "Changes in School Achievement: Project TALENT Ten and Fifteen Year Retests" (Flanagan, 1975) were based. (See Chapter 21, section D.)

POSTSCRIPT

As we explained in the Preface, the History needed a cut-off point, and so the conclusion of activities associated with the 11-year follow-up data collection and processing was selected to serve that function. But events that have occurred since then ought also to be mentioned, if only briefly in order to provide as up-to-date a picture as possible of the current status of Project TALENT.

In the first place, there is the 17-year follow-up. In April of 1976 an agreement was reached between TALENT staff and NIE staff to move up the plans for a 20-year follow-up of TALENT participants and plan instead for a 17-year follow-up. A grant was awarded AIR in November, 1976, to cover the necessary planning activities, including sample selection and instrument development, and associated data bank maintenance activities.

To review the plans a panel was assembled and a meeting was held in New York City in April, 1977. The advisory panel included:

Ronald Abeles
Social Psychologist
Social Science Research Council

Paul Barton
Senior Consultant
National Manpower Institute

Orville G. Brim, President
Foundation for Child Development

Glen Elder
Professor of Sociology
University of North Carolina
On leave at Boystown Center for
the Study of Youth Development

David Featherman*
Professor of Sociology
University of Wisconsin

John Flanagan
Chairman of the Board
American Institutes for Research

Hilda Kahne, Economist
Radcliffe Institute

Herbert Parnes*
Professor of Economics
The Ohio State University

---

*Unable to attend.

Also attending the meeting in an advisory capacity were:

Marion Shaycoft, Principal Research Scientist
American Institutes for Research

Sandra Wilson, Senior Research Scientist
American Institutes for Research

In addition, the following three senior AIR staff members with expertise in areas of interest to the survey were consulted:

Richard Rowe, Vice President
American Institutes for Research

Paul Schwarz, President
American Institutes for Research

Harold Sheppard, Principal Research Scientist
American Institutes for Research

Following this review the plans were amended to incorporate the suggestions of the advisors. A formal request for funds to support the data collection and analysis for the survey, as described in the revised plans, was submitted in the summer of 1977. This request was still pending as of this writing.*

At least one other event of considerable significance to Project TALENT has occurred in 1977. The project's Principal Investigator has "retired." "Retired" is in quotation marks because John Flanagan continues to work essentially full-time and he continues to direct research projects for AIR, on problems in which he is particularly interested. And even though he has relinquished the role of Principal Investigator on Project TALENT, he continues to maintain a lively interest in the project and to act as a consultant to it. We hope Project TALENT will continue for many years to come, and that his interest in it will continue during all those years.

---

* September 1977

REFERENCES

American Institutes for Research. *Planning career goals* (a comprehensive test battery and career guidance system, by John C. Flanagan, Marion F. Shaycoft, & Mary B. Willis). Monterey, Calif.: CTB/McGraw-Hill, 1976.

American Institutes for Research. *The Project Talent Data Bank: A handbook*. Palo Alto, Calif.: Author, 1972.

American Institutes for Research. *The Project Talent Data Bank*. Pittsburgh: Project Talent Office, 1965. ED 072 105

Angoff, W. H. Test reliability and effective test length. *Psychometrika*, 1953, *18*, 1-14.

Brim, O. G., Jr., Goslin, D. A., Glass, D. C., & Goldberg, I. *The use of standardized ability tests in American secondary schools and their impact on students, teachers, and administrators*. New York: Russell Sage Foundation, 1964.

Card, Josefina J. *Consequences of adolescent childbearing for the young parent's future personal and professional life*. Palo Alto, Calif.: American Institutes for Research, 1977.

Claudy, J. G., Gross, D. E., & Strause, Rebecca D. *Two population studies: I. Family size, birth order, and characteristics of young adults, II. A study of married couples in Knox County, Tennessee*. Palo Alto, Calif.: American Institutes for Research, 1974.

Cooley, W. W., & Lohnes, P. R. *Predicting development of young adults*. Palo Alto, Calif.: Project TALENT Office, 1968.

Cooley, W. W., & Lohnes, P. R. *Multivariate procedures for the behavioral sciences*. New York: John Wiley and Sons, Inc., 1962.

Cureton, E. E. *Project TALENT tests as anchors for equating other tests*. Palo Alto, Calif.: Project TALENT Office, 1973.

Cureton, E. E. *A factor analysis of Project TALENT tests and four other test batteries*. Palo Alto: Project TALENT Office, 1968.

Cureton, Louise W. *Arrests for drunkenness of young adults and their parents*. Palo Alto, Calif.: Project TALENT Office, 1973.

Cureton, Louise W. *Early identification of behavior problems*. Washington, D.C.: American Institutes for Research, 1970.

Dailey, J. T., Shaycoft, Marion F., & Orr, D. B. *Calibration of Air Force selection tests to Project TALENT norms*. (Technical Documentary Report PRL-TDR-62-6, 2570th Personnel Research Laboratory [AMD], Lackland Air Force Base, Texas, 1962).

Flanagan, J. C. Changes in school levels of achievement: Project TALENT ten and fifteen year retests. *Educational Researcher*, 1976, *5*(8), 9-12.

Flanagan, J. C. Implications of Project TALENT for research in the teaching of English. In E. Steinberg (Ed.), *Needed research in the teaching of English*. Proceedings of a conference at Carnegie Institute of Technology, Pittsburgh, May 5-7, 1962.

Flanagan, J. C. Maximizing human talents. *The Journal of Teacher Education*, June 1962, *13*(2), 204-215.

Flanagan, J. C. Project TALENT: Preliminary findings. Paper presented at a joint meeting of the American Educational Research Association and the American Association of School Administrators, Atlantic City, February 20, 1962.

Flanagan, J. C. *Flanagan Aptitude Classification Tests technical report*. Chicago: Science Research Associates, 1959.

Flanagan, J. C. *Aptitude classification tests*. Chicago: Science Research Associates, 1953.

Flanagan, J. C. The use of comprehensive rationales in test development. *Educational and Psychological Measurement*, 1951, *11*, 151-155.

Flanagan, J. C., Cooley, W. W., Lohnes, P. R., Schoenfeldt, L. F., Holdeman, R. W., Combs, Janet, & Becker, Susan. *Project TALENT one-year follow-up studies*. Pittsburgh: Project TALENT Office, 1966. 010087

Flanagan, J. C., Dailey, J. T., Shaycoft, Marion F., Gorhan, W. A., Orr, D. B., & Goldberg, I. *Design for a study of American youth*. Boston: Houghton Mifflin, 1962.

Flanagan, J. C., Dailey, J. T., Shaycoft, Marion F., Orr, D. B., & Goldberg, I. *Studies of the American high school*. Washington, D.C.: Project TALENT Office, 1962.

Flanagan, J. C., Dailey, J. T., Shaycoft, Marion F., Gorham, W. A., Orr, D. B., & Goldberg, I. Designing the study. (Technical report to the U.S. Office of Education, Cooperative Research Project No. 635.) Washington, D.C.: University of Pittsburgh, Project TALENT Office, 1960.

Flanagan, J. C. Davis, F. B., Dailey, J. T., Shaycoft, Marion F., Orr, D. B., Goldberg, I., & Neyman, D. A. Jr. *The American high school student*. Washington, D.C.: Project TALENT Office, 1964.

Flanagan, J. C., & Jung, S. M. *Progress in education: A sample survey (1960-1970)*. Palo Alto, Calif.: Project TALENT Office, 1971. 049314

Flanagan, J. C., and Russ-Eft, Darlene. *An empirical study to aid in formulating educational goals*. Palo Alto, Calif.: American Institutes for Research, 1975. 117809

Flanagan, J. C., Shaycoft, Marion F., Richards, J. M., Jr., & Claudy, J. G. *Five years after high school*. Palo Alto, Calif.: Project TALENT Office, 1971. ✓ 053875

Flanagan, J. C., Tiedeman, D. V., Willis, Mary B., & McLaughlin, D. H. *The career data book: Results from Project TALENT's five-year follow-up study*. Palo Alto, Calif.: American Institutes for Research, 1973. ✓ 082569 ⓔ E

Folger, J. K., Astin, Helen S., & Bayer, A. E. *Human resources and higher education*. New York: Russell Sage Foundation, 1970.

Gilmartin, K. J., McLaughlin, D. H., Wise, L. L., & Rossi, R. J. *Development of scientific careers: The high school years*. Palo Alto, Calif.: American Institutes for Research, 1976.

Glaser, R. Instructional technology and the measurement of learning outcomes. *American Psychologist*, 1963, *18*, 518-521.

Goslin, D. A. *The search for ability*. New York: Russell Sage Foundation, 1963.

Jencks, C., *et al*. *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books, In ., 1972.

Jencks, C. S., & Brown, Marsha D. Effects of high schools on their students. *Harvard Educational Review*, August 1975, *45*(3), 273-324.

Kapel, D. E. *Effects of Negro density on student variables and the post-high-school adjustment of male Negroes*. Palo Alto, Calif.: Project TALENT Office, 1968.

Lohnes, P. R. *Measuring adolescent personality*. Pittsburgh: Project TALENT Office, 1966.

Love, W. A., Jr., & Stewart, D. K. *Interpreting canonical correlations: Theory and practice*. Palo Alto, Calif.: Project TALENT Office, 1968.

McLaughlin, D. H., Fulscher, G. V., & Yen, Wendy M. *Project TALENT's special sample: Is it necessary?* Palo Alto, Calif.: American Institutes for Research, 1974.

Rossi, R. J., Bartlett, Wendy B., Campbell, Emily A., Wise L. L., & McLaughlin, D. H. *Using the TALENT profiles in counseling: A supplement to The Career Data Book*. Palo Alto, Calif.: American Institutes for Research, 1975.

Rossi, R. J., Wise, L. L., Williams, Kathleen L., & Carrel, Kathleen L. *Methodology of the Project TALENT 11-year follow-up study*. Palo Alto, Calif.: American Institutes for Research, 1976.

Shaycoft, Marion F. *Guide to the development, evaluation, and use of criterion-referenced tests*, 1976. Note: This monograph, which was prepared under the sponsorship of the U.S. Office of Education (Office of Planning, Budgeting, and Evaluation), is not yet available in published form.

Shaycoft, Marion F. *Careers in science: A Project TALENT study*. Palo Alto, Calif.: American Institutes for Research, 1975.

Shaycoft, Marion F. Development and analysis of a test to estimate size of vocabulary. (Paper presented at the American Psychological Association convention, Chicago, September 7, 1965; revised February 26, 1958) Palo Alto, Calif.: American Institutes for Research, 1968.

Shaycoft, Marion F. *The high school years: Growth in cognitive skills*. Palo Alto, Calif.: Project TALENT Office, 1967.

Shaycoft, Marion F. *Calibration of academic aptitude composite to Project TALENT results*. (Technical report to 6570th Personnel Research Laboratory, Aerospace Medical Division, AFSC, AF 41 (609)-1588, Phase 4.) Washington, D.C.: American Institutes for Research, 1963.

Shaycoft, Marion F. *Comparison of Air Force recruits with segments of the high-school-age population*. (Technical report to 6570th Personnel Research Laboratory, Aerospace Medical Division, AFSC, AF41 (609)-1588, Phases 1-3.) Washington, D.C.: American Institutes for Research, 1963.

Shaycoft, Marion F., Dailey, J. T., Orr, D. B., Neyman, C. A., Jr., & Sherman, S. E. *Studies of a complete age group--Age 15*. Washington, D.C.: Project TALENT Office, 1963.

Shaycoft, Marion F., Neyman, C. A., Jr., & Dailey, J. T. *Comparisons of Navy recruits with male high school students on the basis of Project TALENT data*. (Final report, Department of the Navy, Bureau of Naval Personnel, Nonr-382[00].) Washington, D.C.: American Institutes for Research, 1962.

Shaycoft, Marion F., Wise, L. L., & McLaughlin, D. H. *Minimizing total error: A trade-off problem*. Paper presented at joint session of the American Psychological Association (Division 5) and Psychometric Society, in Chicago, 3 September 1975.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Thorndike, R. L., & Hagen, Elizabeth. *Ten thousand careers*. New York: John Wiley and Sons, Inc., 1959.

Tupes, E. C., & Shaycoft, Marion F. *Normative distributions of AQE aptitude indexes for high-school-age boys*. (Technical Report PRL-TDR-64-17, Personnel Research Laboratory, Aerospace Medical Division, AFSC, Task 771705, Lackland Air Force Base, Texas, 1964.)

Wargo, M. J., Tallmadge, G. K., Michaels, Debbra D., Lipe, D., & Morris, Sarah J. *ESEA Title I: A reanalysis and synthesis of evaluation data* from fiscal year 1965 through 1970. (Final report to the Office of Education for Contract No. OEC-0-71-4766.) (AIR-27400-3/72-FR). Palo Alto, Calif.: American Institutes for Research, 1972.

Wilson, Sandra R., Flanagan, J. C., & Uhlaner, J. E. Quality of life as perceived by 30 year old Army veterans. Arlington, Virginia: U.S. Army Research Institute for the Behavioral and Social Sciences, February 1975.

Wilson, Sandra R., & Wise, L. L. The American citizen: 11 years after high school. Palo Alto, Calif.: American Institutes for Research, 1975.

Wise, L. L., McLaughlin, D. H., & Steel, Lauri. The Project TALENT Data Bank Handbook. Palo Alto, Calif.: American Institutes for Research, 1977.

Yen, Wendy M., & McLaughlin, D. H. Post high school education and career development. Palo Alto, Calif.: American Institutes for Research, 1974.

APPENDIX A1

COMPOSITION OF THE PROJECT TALENT BATTERY

Instruments — Page

APTITUDE AND ACHIEVEMENT TESTS A1-2
  Information Test A1-2
  Other Aptitude and Achievement Tests A1-3

MISCELLANEOUS A1-3

INVENTORIES A1-4
  Student Activities Inventory A1-4
  Interest Inventory A1-4
  Student Information Blank A1-4

APPENDIX C: Composition of the Project TALENT Battery

Variable # Code						
Kind of score*	3 digit #		Options per item	No. of items	No. of minutes working time**	No. of scales****
		APTITUDE AND ACHIEVEMENT TESTS				
		Information Test		(395)	(125)	(43)
		Part I	5	(252)	90	(16)
		Subscales				(15)
R	101	1. Screening		12		1
R	102	2. Vocabulary		21		1
R	103	3. Literature		24		1
R	104	4. Music		13		1
R	105	5. Social studies		24		1
R	106	6. Mathematics		23		1
R	107	7. Physical science		18		1
R	108	8. Biological science		11		1
R	109	9. Scientific attitude		10		1
R	110	10. Aeronautics and space		10		1
R	111	11. Electricity and electronics		20		1
R	112	12. Mechanics		19		1
		a. Tools and construction ***		(10)		
		b. Motors and mechanisms***		(9)		
R	113	13. Farming		12		1
R	114	14. Home economics		21		1
		a. Cooking***		(11)		
		b. Other***		(10)		
R	115	15. Sports		14		1
R	190	Part I Total		(252)		1
		Part II	5	(143)	35	(25)
		Subscales				(22)
R,K	131	1. Art		12		1
R,K	132	2. Law		9		1
R,K	133	3. Health		9		1
R,K	134	4. Engineering		6		1
R,K	135	5. Architecture		6		1
R,K	136	6. Journalism		3		1
R,K	137	7. Foreign travel		5		1
R,K	138	8. Military		7		1
R,K	139	9. Accounting, business, sales		10		1
R,K	140	10. Practical knowledge		4		1
R,K	141	11. Clerical		3		1
R,K	142	12. Bible		15		1
R,K	143	13. Colors		3		1
R,K	144	14. Etiquette		2		1
R,K	145	15. Hunting		5		1
R,K	146	16. Fishing		5		1
R,K	147	17. Outdoor activities (other)		9		1
R,K	148	18. Photography		3		1
R,K	149	19. Games (sedentary)		5		1
R,K	150	20. Theater and ballet		8		1
R,K		a. Theater***		(6)		-
R,K		b. Ballet***		(2)		-
R,K	151	21. Foods		4		1
R,K	152	22. Miscellaneous		10		1
		Overlapping Subscales				(2)
R,K	162	Vocab. (overlapping other scales)		(9)		1
R,K	163	Outdoor activities (Hunting + Fishing + other)		(19)		1
R,K,L	192	Part II Total (inc. 10 misc. items)		(143)		1
		Parts I + II combined		(395)	(125)	(2)
R	172	Vocabulary scale (Variables 102+162)		(30)		1
R	100	Grand Total		(395)		1

Variable # Code						
Kind of score*	3 digit #		Options per item	No. of items	No. of minutes working time**	No. of scales****
		Other aptitude and achievement tests				
R	211	Memory for sentences				
		Study	-	(40 sentences)	6	-
		Test	5	16	10	1
R	212	Memory for words				
		Study	-	(24 words)	2	-
		Practice	-	(24)	2	-
		Test	5	24	4	1
R	220	Disguised words	5	30	3	1
		English		(113)	52	(6)
R	231	1. Spelling	5	16		1
R	232	2. Capitalization	2	33		1
R	233	3. Punctuation				
		a. Punctuation marks***	3-5	(16)		
		b. Sentence structure***	3	(11)		
R		Punctuation Total		27		1
R	234	4. English usage	3-5	25		1
R	235	5. Effective expression	3-5	12		1
R	230	Total		(113)		1
R	240	Word functions in sentences				
		Directions	-	-	2½	
		Test	5	24	15	1
R	250	Reading comprehension	5	48	30	1
R	260	Creativity	5	20	20	1
R	270	Mechanical reasoning	3-5	20	11	1
R	281	Visualization in 2 dimensions	5	24	4	1
R	282	Visualization in 3 dimensions	5	16	9	1
R	290	Abstract reasoning	5	15	11	1
		Mathematics		(54)	50	(6)
R	311	Part I. Arithmetic reasoning	4-5	16		1
R	312	Part II. Introductory h.s. math	5	24		1
R	320	Subtotal (Parts I + II)		(40)		1
R	333	Part III. Advanced h.s. math	5	14		1
R	334	H.S.Subtotal (Parts II + III)		(38)		1
R	340	Total (Parts I + II + III)		(54)		1
R,A,F	410	Arithmetic computation*****	5	72	9	1
R,A,F	420	Table reading*****	5	72	3	1
R,A,F	430	Clerical checking*****	2	74	3	1
R,A,F	440	Object inspection*****	5	40	3	1
		MISCELLANEOUS				
A	500	Preferences test	2	166	3	1
		Themes	-	(2)	(10+)	
-	910	My views about an ideal occupation		1	5+	
-	920	What high school means to me		1	5+	

Variable # Code						
Kind of score*	1 digit #		Options per item	No. of items	No. of minutes working time**	No. of scales****
		INVENTORIES				
	600	Student Activities Inventory	5	(150)	20	
		Regular scales		(108)		(10)
K	601	Sociability		12		1
K	602	Social sensitivity		9		1
K	603	Impulsiveness		9		1
K	604	Vigor		7		1
K	605	Calmness		9		1
K	606	Tidiness		11		1
K	607	Culture		10		1
K	608	Leadership		5		1
K	609	Self-confidence		12		1
K	610	Mature personality		24		1
		Experimental scales		(18)		
	611	Conventionality***		4		
	612	Theoreticality***		8		
	613	Group centeredness***		6		
		Miscellaneous		24		
L	700	Interest Inventory*****	5	205	20	
		Part I. Occupations		(122)		
		Part II. Activities		(83)		
		Scales: ("core items")		(173)		(17)
F*,P	701	Physical science,engineering,math		16		1
F*,P	702	Biological science and medicine		8		1
F*,P	703	Public service		11		1
F*,P	704	Literary-linguistic		16		1
F*,P	705	Social service		12		1
F*,P	706	Artistic		7		1
F*,P	707	Musical		5		1
F*,P	708	Sports		8		1
F*,P	709	Hunting and fishing		3		1
F*,P	710	Business management		14		1
F*,P	711	Sales		6		1
F*,P	712	Computation		10		1
F*,P	713	Office work		7		1
F*,P	714	Mechanical-technical		15		1
F*,P	715	Skilled trades		18		1
F*,P	716	Farming		7		1
F*,P	717	Labor		10		1
		Miscellaneous: "non-core items"		(32)		
L	800	Student Information Blank*****	2-36	(394)	80	(18)
P*	801	Socioeconomic index		9		1
D	802	H.S. curriculum: acad. vs other		1		1
F	803	H.S. courses taken (acad. orient.)		7		1
P	820	H.S. grades		7		1
F	822	Guid. received in h.s.		10		1
F	823	Guid. received elsewhere		6		1
P	827	Study habits and attitudes		16		1
P	828	Self-perception of writing skills		4		1
P	829	Self-perception of reading skills		6		1
F	830	Amount of extracurricular reading		5		1
P	831	Var. of extracurr. group activities (except sports)		10		1
F	832	Degree of participation in extracurr. group activities (except sports)		10		1
P	833	Variety of hobbies (except sports)		15		1
F	834	Degree of act. in hobbies (exc. sports)		15		1
F	835	Participation in sports		4		1
F	836	Leadership roles		3		1
F	837	Social life		5		1
F	838	Work activities (chores & jobs)		5		1

* The "kind of score" column shows the kinds of scores routinely available for Project TALENT cases in general. The meaning of these code letters is explained in Appendix B-2. Certain additional scores may be available for certain cases, such as the "4% sample" (Subsample A-04.0-25) described in The American High School Student (Flanagan et al., 1964, Chapter 2).

** Does not include the time used for giving directions except where otherwise indicated. (The exceptions occur where comprehension of directions is considered an integral part of the testing time allowance.)

*** In the regular TALENT testing (1960), scores on this variable are available for the "4% sample" (Subsample A-04.0-25) only.

**** The term "scale," as used in this table, means set of items. The numbers appearing in the "No. of Scales" column represent the number of scales, including composites, for which scores are available routinely for the Project TALENT cases in general.

***** For information about scoring procedures on these variables see Appendix E.

Appendix A2. THUMBNAIL DESCRIPTIONS OF THE TESTS

Variable Number

Information Test: Part I

This test contains 252 items which are divided into 15 subscales, as follows:

101 Screening. This scale contains questions which are extremely simple and are basic knowledge to all elementary school children. It was designed to help in identifying mentally retarded students, others who are functionally illiterate, and those who took the test with a flippant or apathetic attitude.

102 Vocabulary. This score gives some indication of the relative size of the student's general vocabulary.

103 Literature. The purpose of this score is to measure familiarity with the world of literature. Both prose and poetry are included. Some, but not all, of the literary works on which test items are based are required reading in many schools, and they are on recommended reading lists in other schools. However, the items are not limited to reading that is likely to have been done in school. The test's broad coverage makes it likely that a student who has acquired the habit of recreational reading will get a reasonably good score regardless of what specific books are required reading in his school.

104 Music. This score is intended to indicate amount of musical information, not musical talent. Those who enjoy going to concerts and opera, or listening to serious music on radio and phonograph, are as likely to get a good score as those having formal training in music.

105 Social Studies. This scale covers facts and concepts from the fields of history, economics, government and civics, geography, and current affairs.

Variable Number

106 Mathematics. These items are concerned with definitions, the vocabulary of mathematics, mathematical notation, other kinds of factual information, and the understanding of mathematical concepts. None of the items requires computation or reasoning or calls for solving a problem, since these abilities are covered by other tests in the battery (the Arithmetic Computation Test and the Mathematics Test).

107 Physical Sciences. This scale includes items about chemistry, physics, astronomy, and other physical sciences. Many of the items cover information that might readily be acquired in other ways than through formal instruction.

108 Biological Sciences. This scale includes items about botany, zoology, and microbiology. A few items about nature lore are included, though most of the items are concerned with more formal aspects of biological science.

109 Scientific Attitude Scale. These items provide a subscore which should be indicative of how the individual views the world--whether he views it as a place where there are logical cause-and-effect relationships, or whether he regards it as a place where consequences are illogical and arbitrary. In the latter category fall the modern-day equivalents of the primitive beliefs that have been called "sympathetic magic". Also in the category of illogical and arbitrary consequences are the premises of palmistry, astrology, and numerology.

Each item consists of a description of a phenomenon or an occurrence. Five explanations are presented, and the task is to select the "best" one. Only one of the five explanations is reasonable and logical. Of the other four, some involve common superstitions, magic, concepts incompatible with the scientific viewpoint, or belief in the occult. Selection of other distractors may primarily imply muddled thinking on the part of the examinee.

Variable
Number

0 Aeronautics and Space. These items are on such topics as flying technique, navigation, jet planes, and space exploration. Much of the information that the student has in this area is likely to have been acquired out of school.

111 Electricity and Electronics. These items stress information that is acquirable through direct experience in the construction and maintenance of electrical and electronic equipment. Students who have worked on radios, hi-fi sets, or other electronic equipment, or on mechanisms with electric motors, should get good scores. Since the range of content covered is fairly broad, it seems likely that the average high school student will be able to answer many of the questions if he is at all interested in the area.

112 Mechanics. These items tap a wide range of information. Many of them are concerned with automobiles; others are concerned with other common machines and tools with which people who are interested in mechanical activities are likely to be familiar. The emphasis is on information that is likely to be acquired through direct experience with tools, engines, and motors. The scores should also be influenced by the amount of experience and training in mechanics the student has had.

113 Farming. These items are intended to give some representation to the information that children who grow up on farms or ranches are likely to acquire. The items might give some advantage to those who have an interest in rural life.

114 Home Economics. These items test information on cooking, sewing, caring for babies, cleaning, and other activities of a domestic nature. Students who have engaged in such activities, whether as a chore, or part-time job (e.g., babysitting), or just because they enjoy them, as well as students who have had formal instruction in home economics, should have an advantage on these items.

Variable Number

115 Sports. These items are intended to measure general familiarity with a wide range of sports and sports terms. Knowledge of specific sports figures is not required.

190 Information Test: Part I Total. The 15 scales described above are combined to give a Part I total score.

Information Test: Part II

This test consists of an additional 143 items which can be divided into 23 subscales mainly testing information acquired out of school. These subscales are numbered 131 through 152, and 162. Scale 162 (Vocabulary) partially overlaps some of the other scales.

Information Test: Parts I and II Combined

172 Total Vocabulary. Vocabulary scales from Part I and Part II are combined to yield an overall vocabulary score.

100 Information Test Total. The total score, which represents breadth of general information, is based on the 395 items in Part I and Part II. Total score on an information test may be regarded as one very good indicator of general ability to learn, something called "intelligence", at least in the case of students whose cultural and educational backgrounds have been normal. However, students from extremely deprived backgrounds, who have not had average opportunities to acquire a broad range of information, will probably have more general ability than is indicated by their score on this test.

211 Memory for Sentences.

This score indicates one kind of memory--the ability to memorize simple descriptive statements and recall a missing word when the rest of the sentence is provided sometime later. It is important to recognize that only one very specific kind of memory is tested, and to avoid assuming that the score indicates how well the student can do in all kinds of memory tests.

Variable Number

212 Memory for Words.

The purpose of this test is to measure another kind of rote memory--the ability to memorize foreign words corresponding to common English words. This ability is obviously directly relevant to the learning of a foreign language. It is also presumably related to the ability to learn many other kinds of material.

220 Disguised Words

This test is designed to measure the ability to form connections between letters and sounds. This is believed to be related to the ability to puzzle out from context and appearance the meaning of a word which is probably reminiscent of a familiar English word. This ability is probably one aspect of aptitude for learning shorthand or a foreign language.

English Test

The purpose of this test is to measure ability to express oneself adequately in English. A total score and five separate subscores are reported: spelling, capitalization, punctuation, English usage, and effective expression. The test is primarily concerned with written English, but presumably some generalization to spoken English on the basis of the English Usage and Effective Expression subscores is justifiable.

231 Spelling. This score is intended to indicate ability to spell--not size of vocabulary. Students who do not have especially large vocabularies but who are able to spell most of the words they have encountered should get good scores.

232 Capitalization. This score indicates degree of mastery of the rules of capitalization.

233 Punctuation. The purpose of this test is to measure knowledge of the appropriate use of all standard punctuation marks. Considerable emphasis is placed on whether the student has mastered the concept of what constitutes a sentence.

Variable
Number

234 English Usage. This score measures knowledge of preferred usage.

[illegible]5 Effective Expression. This score is intended to measure recognition of good prose expression--in other words, recognition of whether an idea has been expressed clearly, concisely, and smoothly.

230 English Total. This score is the sum of the five subscores described above. It indicates overall achievement in various aspects of English expression which can be measured by objective test items.

240 Word Functions in Sentences.

This test is intended to measure the student's sensitivity to grammatical structure. The fact that the terminology of grammar is not used at all in the test helps reduce the effects of formal training to a minimum. To score well, one must understand something about structure of a sentence and recognize the function of each word or phrase in the sentence. This ability is probably also related to the ability to learn the formal rules of English grammar. The test was made difficult deliberately, since there is reason to believe that a hard test would predict ability to learn foreign languages better than an easier one.

250 Reading Comprehension

The purpose of this test is to measure the ability to comprehend written materials. The ability measured is the ability to read with comprehension, rather than mere ability to mouth or recognize the printed word without understanding the fact, idea, or concept that the writer is attempting to convey. The test includes passages on a wide range of topics. The student reads the passage and then answers a number of questions about it, referring back to the passage as often as he likes. None of the items is answerable without reading the passage. The ability measured by this test is a good predictor of school success in an academic or liberal arts curriculum. In the unusual situation where a student scores low on the reading comprehension test, but obtains good scores on other types of intellectual tests, he may have a specific reading disability that can be corrected by special remedial training.

Variable Number

260 Creativity

The purpose of this test is to measure the ability to find ingenious solutions to a variety of practical problems. High scores on this test should be interpreted as indicating inventiveness or creative ingenuity.

270 Mechanical Reasoning

The purpose of this test is to measure the ability to visualize the effects of the operation of everyday physical forces (such as gravitation) and basic kinds of mechanisms (for instance gears, pulleys, wheels, springs, levers). A kind of reasoning which is related to mechanical aptitude is involved. Although all of the items can be answered without experience in woodworking or other crafts, or in working with motors, past training and experience must nevertheless be borne in mind in interpreting the results.

290 Abstract Reasoning

This is a non-verbal test to measure one kind of abstract reasoning ability--the ability to determine a logical relationship or progression among the elements of a complex pattern, and to apply this relationship to identify an element that belongs in a specified position in the pattern. Some youngsters who have had little schooling or who have had serious cultural or environmental handicaps which result in low scores in vocabulary and reading comprehension may make a high score on the abstract reasoning test, indicating greater potential for academic work then has been developed. Among typical students, who have had a normal environmental background, the abstract reasoning score should be considered just another element in the general domain of intellectual ability.

Visualization Tests

The general purpose of these two tests is to measure spatial visualization. The two tests are described briefly below.

281 Visualization in Two Dimensions. This test measures the ability to visualize how diagrams would look after being turned around on a flat surface, in contrast with the way they would look after being turned over.

Variable
Number

282 Visualization in Three Dimensions. This test measures the ability to visualize how a figure would look after manipulation in three-dimensional space (more specifically, how a two-dimensional figure would look after it had been folded to make a three-dimensional figure).

Mathematics Test

The mathematics Test has three parts.

311 Mathematics Part I (Arithmetic Reasoning). This test is designed to measure the ability to do the kind of reasoning required to solve arithmetic problems. Computation, except at the very simplest level, is excluded from the test.

312 Mathematics Part II (Introductory High School Math). The purpose of this subtest is to measure achievement in all kinds of mathematics generally taught up to and including 9th grade, with the exception of the areas covered in the Arithmetic Computation Test and in Mathematics Part I (Arithmetic Reasoning). The primary emphasis of this test is on elementary algebra; other topics include fractions, decimals, percents, square roots, intuitive geometry, and elementary measurement formulas. While the topics covered are taught in Grade 9 or earlier in most schools, curricula differ considerably in regard to grade placement of various topics.

333 Mathematics Part III (Advanced High School Math). This subtest covers topics normally taught in Grades 10-12 in college-preparatory courses. The items are intended primarily to test understanding and application of basic concepts and methods, not rote memory. A wide range of subjects is included: plane geometry, solid geometry, algebra, trigonometry, elements of analytic geometry, and introductory calculus. It should be noted that some of these subjects are not offered in most high schools. However, students who have taken college preparatory mathematics beyond Grade 9 level, and have really understood it, should be able to score well.

Variable Number

340 Mathematics Total (I + II + III). This is the total mathematics score. It is a comprehensive measure of achievement in mathematics.

Tests of Clerical and Perceptual Speed and Accuracy

410 Arithmetic Computation. The purpose of this test is to measure speed and accuracy of computation. The test is limited to the four basic operations (addition, subtraction, multiplication, and division), and to whole numbers. The aptitude measured is an important one. However, a student does not necessarily need an outstandingly good score on this test in order to do well in mathematics in high school and college.

420 Table Reading. The purpose of this test is to measure speed and accuracy in a non-computational clerical task, involving obtaining information from tables. This kind of clerical aptitude is somewhat more complex than that measured by the Clerical Checking Test described below.

430 Clerical Checking. This test is designed to measure speed and accuracy of perception in a very simple clerical task. The test involves comparing pairs of names to determine whether they are identical.

440 Object Inspection. The purpose of this test is to measure speed and accuracy in perception of form. More specifically, the test is intended to measure the ability to spot differences in small objects quickly and accurately when comparing them visually.

Other

500 Preferences Test. One purpose of this test--frankly an experimental one--is to measure the speed with which a person can reach at least one kind of decision. The test, with a time limit of three minutes, contains 166 items, each consisting of a pair of adjectives describing a person. The student is required to select the adjective from each pair that he would prefer to have applicable to his friends. The two characteristics are intended to be about equal in social acceptability.

In considering the results derived from this test, bear in mind that the test was still in the experimental stage of development when given. We must remember, too, that only one limited kind of decision-making is being tested. This trait may, or may not, be important. It is hypothesized that the test locates the individual on a continuum that has "snap judgments" at one end and "indecision" at the other.

Appendix B. CHOOSING WHICH 1960 VARIABLES TO USE

A. Some general advice

In planning a study using TALENT data it is necessary to decide which of the vast number of 1960 variables to use, since it is seldom either feasible or desirable to incorporate all of them in a single study. In studies in which (whether for reasons of cost or for some theoretical reason) it is only feasible to incorporate a very limited number of variables, the choice of precisely which ones to include become crucial. And in making a decision of this sort we cannot emphasize too strongly that it is inadvisable to depend on instrument name or scale name. For instance the score on the "Preferences Test" is not intended to tell anything about the student's preferences; rather it tells us something about how fast he makes decisions (in a certain limited area). But we could not name it the "Speed-of-Decision-Making Test"; that would wreck it completely by revealing its purpose to the examinees. Similarly, the Student Activities Inventory does not tell, and was not intended to tell, anything about what activities the student engages in. "Student Activities Inventory" was selected as a cover name for a personality inventory.

Of course it is possible to read the thumbnail descriptions of the tests that appear in Appendix A2, and find out the facts indicated above, but thumbnail descriptions can go only so far. For many (though not all) purposes it is important that the researchers have a more exact knowledge of the character of the instrument--precisely what kinds of items make it up, what areas are covered, etc.

Researchers who need more information than the instrument names or scale names provide (and virtually all researchers using TALENT data fall in this category) are referred to the thumbnail descriptions in Appendix A2. Those who need more information than that appendix provides (and many will) are referred to "Design for a Study of American Youth" (Flanagan, Dailey, Shaycoft, Gorham, Orr, and Goldberg, 1962). And those who require still more information about the tests, inventories, and questionnaires are urged to look at the instruments themselves. Specimen Sets of the TALENT Battery are available for this purpose.* For many purposes they should also familiarize themselves with the reliability of the tests. The best currently available source of this information (at least for those variables included in it) is "The High School Years: Growth in Cognitive Skills" (Shaycoft, 1967), table 4-7, columns 5 and 6, in which grade 12 reliabilities are shown. (The corrected-for-range procedure may be applied in order to obtain reliability estimates applicable to other grades or subgroups.)

In the above paragraphs we have urged researchers to recognize the importance of making a wise choice of variables to be included. In making such a choice, some researchers may find it useful to base their decisions in part on suggestions derived from past research results and extensive use of the TALENT variables. Providing such suggestions is the intended purpose of the paragraphs that follow.

B. Some more specific suggestions

1. Variables with broad-spectrum usefulness

Among the most useful general-purpose or multiple-purpose variables are the following background and test variables.

Background variables

P*801 Socioeconomic index

---

* Specimen Sets may be ordered by writing to:

American Institutes
for Research
P. O. Box 1113
Palo Alto, CA 94302

The current price is $3.00.

Test variables

R-190	Information Part I Total*
R-100	Information Test I Total*
R-172	Vocabulary Total
R-230	English Total
R-250	Reading Comprehension Total
R-260	Creativity
R-270	Mechanical Reasoning
R-280	Visualization in Three Dimensions
R-290	Abstract Reasoning
R-311	Arithmetic Reasoning
R-312	Introductory High School Math

and to a somewhat lesser degree:

F-410	Arithmetic Computation

A priori test score composites

C-002	General Academic Aptitude Composite
C'004	Math Composite
C'005	Technical Information Composite

2. Using the above list

The background variable singled out for listing above, Socioeconomic Index (P*801) has proven useful mainly as a control variable or covariate (a variable where effect is to be partialed out, through the use of partial or part correlations, or covariance analysis).

The test variables listed in section B1 above are the ones that measure abilities and kinds of information acquisition that are useful in a wide variety of contexts. Researchers who for one reason or another must limit their studies to a comparatively small number of test variables might do well to make their selection primarily from those listed above, supplemented, perhaps, by a small number of appropriate variables from the Information Test.

It should be noted that in the list of test variables above, Arithmetic Computation (F-410) is given a somewhat restricted endorsement. This is because its unknown (though probably fairly high) reliability may limit its usefulness in certain contexts.

---

* One or the other of these two variables (R-190 and R-100) might be used in a study, but it would usually be undesirable to use both in the same study (excpet for the purpose of obtaining empirical data as to which of the two would be preferable for the particular purpose being investigated).

C. Answers to specific questions about cognitive variables

Sometimes researchers who wish to use a fairly small number of cognitive variables, perhaps five to ten, find themselves bewildered by the almost astronomical number available. On the basis of our familiarity with the variables--how they behave and what their properties are--we can offer some advice in situations of this sort. Here are some of the questions that might be asked, together with the answers that represent the best advice we can give.

Question 1. Should I use individual variables, such as

R-250	Reading Comprehension
R-290	Abstract Reasoning
R-311	Arithmetic Reasoning
R-312	Introductory High School Math
R-230	English
R-112	Mechanical Information
R-270	Mechanical Reasoning

or should I use *a priori* composites, such as:

C-001	I.Q. Composite
C-004	Quantitative Aptitude Composite
C-003	Verbal Composite
C'005	Technical Information Composite
C-005	Technical Aptitude Composite
C-006	Scientific Aptitude Composite
C-002	General Academic Aptitude Composite

or should I use some of each?

Answer: If you are going to use five or ten variables, all of them should be selected from among the individual variables. To the extent you use composite variables either in conjunction with other composites, or in conjunction with individual variables, you would be artificially and unnecessarily introducing overlapping of variables into the battery with possibly misleading results. For instance the intercorrelations among the variables would be spuriously high. The only exception to this might be if you wanted to use several individual variables and in conjunction with them one or two composites that do not overlap either each other nor any of the individual variables. An example of such a set of variables* would be:

R-250	Reading Comprehension
R-230	English
R-260	Creativity
R-290	Abstract Reasoning
R-109	Scientific Attitude
R-270	Mechanical Reasoning
C'004	Math Composite
C'005	Technical Information Composite

One important point in conjunction with all of the advice offered above is that it is intended to be applicable to substantive research only. Of course, if you want to do some sort of methodological research involving an empirical study of the effects of overlapping on the correlations of composites,** you would want to incorporate several overlapping composites in the set of variables in your study.

Question 2. If my research design permits only one cognitive variable and the question the research is supposed to answer is such that a variable representing "general academic ability" or "overall ability" or something equally nonspecific is needed, what one variable would you recommend?

Answer: Variable C-002 (General Academic Aptitude Composite) is an excellent variable for the purpose. It is an overall ability composite incorporating many different kinds and aspects of ability, and it yields extremely reliable scores.

---

* Readers: Please do not copy this set of variables to use for your research, unless you have some reason for thinking it is appropriate for your particular study. It is presented here merely as an example of a nonoverlapping set involving composites, and not because it is believed to have any universal merit for research studies in general.

** But you probably wouldn't want to study that particular question since the answer to it is already known.

Question 3. If I want some sort of general information score, which of the following three is best?

R-190	Information Part I Total
R-192	Information Part II Total
R-100	Information Total (which equals R-190 plus R-192)

Answer: For most purposes, R-100, being the most general, is best. Actually the other two composite information scores, R-190 and R-192, were computed largely as a matter of practical convenience, rather than because there was any particular rationale for them. In organizing the battery in terms of test booklets and answer sheets, it had been necessary to put part of the Information test on one of the answer sheets that was to be scored immediately so that reports of results could be sent to the participating schools. But since there wasn't room on the answer sheet for the entire information test, it had to be split in two, with the second portion put on one of the answer sheets scheduled for later processing. However because the first part was to contain the subscores that were to be reported to the schools, most of the directly school-related topics are in Part I of the Information Test, while Part II contains more of the items that are more likely to be learned outside of an academic program in high school than *in* such a program. What this means is that while R-100 (the Information Total) is probably the best all-purpose general information score, for a ready-made score that emphasizes (but is not based solely on) school-taught subjects, R-190 is useful; and R-192 likewise may provide a useful score on general information probably acquired elsewhere than in formal education.

Question 4. Which of the following three vocabulary scores is best to use?

R-102	Vocabulary I
R-162	Vocabulary II
R-172	Vocabulary Total (equals R-102 plus R-162)

Answer: The answer to this question is almost, but not quite, parallel to the answer to Question 3 above. Since Vocabulary I is a subset of the Information Part I items and since Vocabulary II is a subset of the Information Part II items, the Vocabulary Total score (R-172) is the best overall vocabulary score. However, there is a special consideration that is applicable to R-162 (and consequently to R-172) that is not applicable to any other subscore routinely yielded by the Information Test; in Part II of the Information Test, none of the items of the vocabulary scale, R-162, is in that scale exclusively; instead, each such item is jointly in the Vocabulary Scale and in one other Part II subscale. Thus if the nature of the data analysis to be carried out is such that overlapping scales should not be included, and if any of the Information Part II scales that include items also in Vocabulary II are to be included, then clearly the only vocabulary scale that should be seriously considered is Vocabulary I (R-102).

Question 5. If I want to include one or more measures of English ability should I use the English Total score (R-230) or some or all of the following part scores,

R-231	Spelling
R-232	Capitalization
R-233	Punctuation
R-234	English Usage
R-235	Effective Expression

or both?

Answer: It depends on what the purpose of your study is. For most purposes using just the English Total will be entirely adequate. But if you are concerned with the question of predicting ability in a specific occupation or group of occupations in which different aspects of English are specifically important in their own right, rather than just as contributing to overall ability in English, you will probably want to use some or all of the part scores.

Suppose, for instance, that the occupation you are interested in is secretary; or suppose it is book editor. For many secretarial jobs and for almost all editing jobs, ability to spell well is important--and spelling does not have a particularly high correlation with the other components of the English test.

Remember, however, that if you are planning to compute a correlation matrix and then use it in some sort of multivariate analysis (under which heading we include multiple correlation, factor analysis, components analysis, multiple discriminant analysis, canonical correlations, etc.) you will not want to include both the total score and all its components in the same multivariate analysis.

Question 6. There are a lot of different mathematics variables. Which should I use?

Answer: A wide variety of data analyses that have been carried out involving both Math Information (R-106) and Introductory High School Math (R-312) suggest that for reasons that are not crystal-clear, the two measures have essentially the same factorial composition even though they were constructed to different specifications. Thus if the number of test variables you want to include is strictly limited, you might want to include one of these two variables, but almost certainly not both.

If you wish to limit your consideration to the kind of math that is involved in knowing how to solve "word problems" that can be solved without the use of algebra, use the Math I score (Arithmetic Reasoning). If you are concerned about the level of mathematical skills acquired by students (especially upperclassmen) who have taken or who might have taken math beyond grade 9, use the Math III score (Advanced High School Math).

Question 7. If I want to include some sort of a math composite or overall math score should I include the "Math Total" score (R-340), which is the raw score total of R-311, R-312, and R-333, or should I use "Math Composite" score (C'004) which is a weighted sum of the same three part scores that go into R-340, plus the "Math Information" score (R-106)?

Answer: For reasons explained in the answer to Question 6 above, it doesn't make a lot of difference whether you use R-340 or C'004. The latter may be slightly preferable because it is based on more items and is therefore a bit more reliable. But both have excellent reliability. It should be noted that in the opinion of the author, for most purposes the C'004 math composite is preferable (by a microscopic amount) to C-004 (the Quantitative Aptitude Composite), which has exactly the same components but with slightly different weights.

Question 8. There are two tests called memory tests:

R-211 Memory for Sentences
R-212 Memory for Words

If I want to use only one memory test, which one should I use?

Answer: Both of these are tests of memorization ability, and the question as to which one is preferable is in this instance easy and unambiguous. By all means use Memory for Words (R-212). The Memory for Sentences total was included in the battery for experimental purposes only, as indicated in Chapter 5 section C. Although the nature of the test is such that its reliability cannot be determined, there is good reason to suppose its reliability is very low. Perhaps for that reason, it has not been found to have predictive value for anything at all. The Memory for Words test, on the other hand, is parallel to a test that has proved its usefulness in Carroll's Psi-Lambda Battery, as a measure of aptitude for learning of foreign language, and has also proved to have predictive value in other contexts. Although this test, like Memory for Sentences, is one for which a reliability coefficient cannot be computed, the test's correlations with some other variables are sufficiently high to warrant the inference that it has good reliability.

Question 9. There are also two "visualization" tests:

R-281 Visualization in Two Dimensions
R-282 Visualization in Three Dimensions

Which is the better one if just one is to be used?

Answer: These two tests appear to measure somewhat different functions. Besides the difference in dimensionality, the two-dimensional test (R-281) is considerably more highly speeded than the three-dimensional test (R-282). In general, the Visualization in Three Dimensions test is the one that has turned out to have better predictive value. Thus if you feel you are limited to one visualization test, R-282 is generally to be preferred to R-281. However, it should be borne in mind that it might be worthwhile to use both.

Question 10. I want to measure mechanical ability. There are two variables with the word "mechanical" in their names:

R-112	Mechanical Information
R-270	Mechanical Reasoning

Which should I use?

Answer: Use both if possible. These tests are both useful, and they differ substantially. Their names happen to be quite descriptive. R-112 measures information, while R-270 measures ability to understand and reason about mechanical relations, without requiring any substantial amount of specific technical information.

D. Noncognitive variables

The question sometimes arises as to which set of noncognitive measures--the Interest Inventory scores or the scores on personality traits yielded by the SAI--is more useful, when one doesn't want to use both. Here the answer is unambiguously in favor of the Interest scores. Although we do not yet have reliability data on either of these instruments, and therefore our conclusions have to be a little more guarded than if we did, all the correlational evidence now available suggests that of the ten regular scales in the SAI only the impulsiveness scale has any substantial degree of uniqueness.

The other nine SAI scales all seem to be measuring one common factor (which is probably best represented by the Mature Personality scale, if one wishes to choose a single scale from the SAI). If the design of one's research study allows room for two SAI variables, the author of this History suggests that for most purposes the best ones to include might be Mature Personality and Impulsiveness.

GLOSSARY OF ACRONYMS AND ABBREVIATIONS

AERA	American Educational Research Association
AFOQT	Air Force Officer Qualification Test
AFPTRC	Air Force Personnel and Training Research Center
AFQT	Armed Forces Qualification Test
AIR	American Institutes for Research
APA	American Psychological Association
AQE	Airman Qualifying Exam
ARI	Army Research Institute
BAVTE	Bureau of Adult, Vocational, and Technical Education, O.E.
BEH	Bureau of Education of the Handicapped, Office of Education
BEPD	Bureau of Educational Personnel Development, O.E.
BESE	Bureau of Elementary and Secondary Education, O.E.
BHE	Bureau of Higher Education, Office of Education
BuPers	Bureau of Naval Personnel
CTB	California Test Bureau
CTS	Cooperative Test Service
DAT	Differential Aptitude Tests
ETS	Educational Testing Service
FAA	Federal Aviation Administration
FACT	Flanagan Aptitude Classification Tests
FY 71	Fiscal Year 1971 (7/1/70 - 6/30/71)
FY 72	Fiscal Year 1972 (7/1/71 - 6/30/72)
GATB	General Aptitude Test Battery
GW	George Washington University

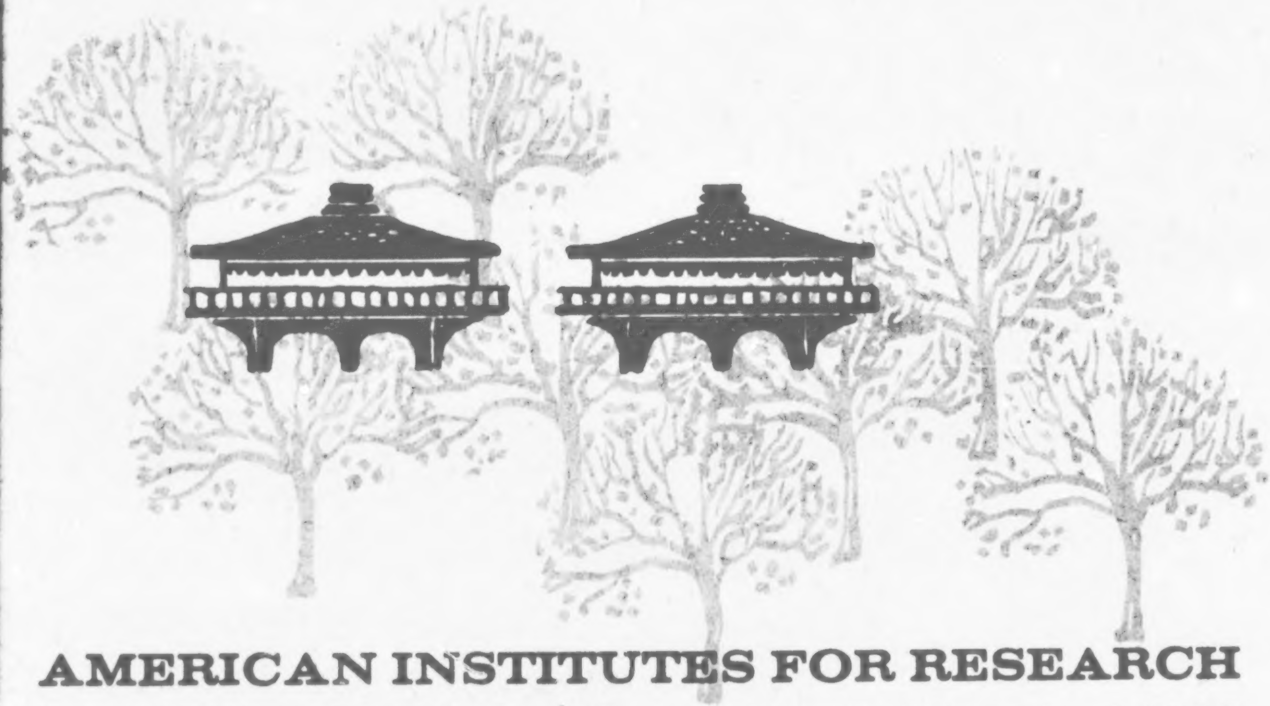
HEW	Department of Health, Education, and Welfare
HIBK	Had I But Known...! (see Preface)
IRE	Institute for Research in Education (one of AIR's institutes)
KR	Kuder-Richardson (a kind of reliability coefficient)
MAP	Measuring Adolescent Personality
MRC	Measurement Research Center
NCERD	National Center for Educational Research and Development
NCES	National Center for Educational Statistics
NCS	National Computer Systems
NIAAA	National Institute on Alcohol Abuse and Alcoholism
NIE	National Institute of Education
NIH	National Institutes of Health
NIMH	National Institute of Mental Health
NSF	National Science Foundation
OE	Office of Education
ONR	Office of Naval Research
PCG	Planning Career Goals
Pitt	University of Pittsburgh
PLAN	Program for Learning According to Needs
PRB	Personnel Research Branch
SAI	Student Activities Inventory
SES	Socioeconomic status index
SIB	Student Information Blank
UC	University of California
TALENT	Not an acronym!
T.E.A.R.S.	Test Editing and Reformatting System

AIR-FR-55400-9/77

Project TALENT: A Short History of a L--o--n--g Project

Technical Appendices

MARION F. SHAYCOFT



AMERICAN INSTITUTES FOR RESEARCH
Post Office Box 1113 / Palo Alto, California 94302

AIR-FR-55400-9/77

Project TALENT:
A Short History of a L--o--n--g Project

Technical Appendices

MARION F. SHAYCOFT

This work was done under Grant No. NIE-G-74-0003 from the National Institute of Education, Department of Health, Education, and Welfare. The content does not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

AMERICAN INSTITUTES FOR RESEARCH
PALO ALTO, CALIFORNIA

1977

TECHNICAL APPENDICES

Page

Appendix C. The Coefficient of Internal Precision: A New Psychometric Statistic for Heritability Research . . . . C- 1

Appendix D1. Must All Members of a Probability Sample Be Randomly Selected? D1- 1

Appendix D2. The "Closed School-Cluster" Method of Selecting a Probability Sample D2- 1

Appendix E1. When is Bias Better? E1- 1

Appendix E2. A Stratified Procedure to Compensate for Nonrespondents E2- 1

Appendix E3. Minimizing Total Error: A Trade-Off Problem E3- 1

Appendix F. Statistical Procedure for Inferring Score Distribution of Unselected Group of 18-Year-Olds from Score Distributions of Unselected Group of 15-Year-Olds and Twelfth-Grade Students . . . . F- 1

Appendix G1. A New Multivariate Index for Use in Educational Planning G1- 1

Propinquity Analysis: A New Guidance Tool . G1-20

Appendix G2. The Project TALENT Approach to the Career-Grouping Problem: Cutting the Gordian Knot G2-1

8/12/75

Appendix C

The Coefficient of Internal Precision: A New Psychometric Statistic for Heritability Research*

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

A. The problem.

One of the things that must be frustrating to people who have been trying to compute heritability indexes is that the most reasonable procedures often give the most unreasonable results. For instance it is not at all unusual to discover, through the use of a perfectly reasonable formula for heritability ratio, that over 100 percent of the variance on some test is due to heredity. That doesn't leave much for environment!

If results of this sort occurred only very infrequently, they might quite reasonably be attributed to freakish sampling errors. But unreasonably high heritability ratios occur so often that the suspicion arises that some kind of systematic error has gotten into the procedure. And, as things turn out, the suspicion often proves justified. The error lies with the reliability coefficients used in the heritability index to correct for unreliable variance.

---

*This is a slightly revised version of a paper presented August 31, 1968, at American Psychological Association convention, San Francisco, as part of a symposium on "Methodological Considerations in Determining the Relative Roles of Heredity and Environment."

What often happens is that twins are given the same form of the same test and then an effort is made to infer percentage of variance due to heredity and percentage due to environment,on the basis of a comparison of the correlation between monozygotic twins' scores and the corresponding correlation for dizygotic twins. The heritability ratio may be defined as the ratio of variance due to heredity to total reliable variance. This necessitates something equivalent to correcting the correlation coefficients for attenuation — and that is where the question of suitable reliability coefficients enters the picture. Using a gross underestimate of the reliability of the variable will make the denominator of the heritability ratio too low and will therefore cause the ratio itself to be too high.

The problem is that there are many different sources of "error variance" or "unreliability" — and that one of these sources, which enters into almost all the standard kinds of reliability coefficients, is inappropriate whenever the correlation being corrected for attenuation is between scores based on the same form OF THE TEST. The inappropriate source of error variance is variance due to the specific sample of test items — and it is obvious why it is inappropriate when the correlation between twins is based on both twins' having scores on the same set of items. If a conventional reliability coefficient, which is reduced by the effect of variation between sets of items, is used to evaluate a correlation coefficient not affected by this source of variation, the correlation corrected for attenuation — in other words the correlation that would be obtained between perfectly reliable measures — will appear to be higher than it really is.

Parallel forms reliability coefficients, split-half coefficients, and the various Kuder-Richardson formulas all reflect, to some extent, the effects of error variance due to the specific set of items. The only ordinary kind of reliability coefficient not thus affected is retest reliability. And because of at least two serious limitations, there are very few kinds of tests for which retest reliability is appropriate. Because of the large spurious inflation due to memory, retest reliability is ruled out as a procedure, for virtually all tests in the cognitive domain. Practice effects constitute the second major limitation. These are not restricted to cognitive tests, but can affect some tests in the psychomotor domain, perceptual speed-and-accuracy tests of various kinds, etc. Since practice effects are not identical for all persons, the retest reliability coefficient is misleadingly affected by them.

These limitations mean that for most tests retest reliability can't be used. And when a single form of a test is used for both twins of a pair all the other methods -- parallel forms, split-half, Kuder-Richardson — are also ruled out. Of course one solution might be to use parallel forms for the twins. But there are a lot of reasons why this is usually a less-than-perfect solution to the problem. One very practical reason is that for many tests no parallel forms are available. Another difficulty is that when parallel forms _are_ allegedly available there are certain difficulties in ascertaining whether two forms are appropriately equivalent, inadequately equivalent, or excessively and spuriously equivalent (as they would be if they were virtually identical). Furthermore use of different forms for the two twins of a pair may tend to muddy the results a bit and make interpretation somewhat more difficult, even when the two forms are genuinely and appropriately equivalent.

B. Nature of the coefficient of internal precision

1. Kinds of error variance included.

But if a single form of the test is to be used for twins and if no suitable kind of reliability coefficient is available to use in interpreting the resulting correlation between twins, it appears necessary to invent a new statistic, to fill the gap. The coefficient of internal precision, represented by a lower-case gamma, is such a statistic. It isn't a reliability coefficient but it is something like one. Unlike a reliability coefficient, it is obtained on the basis of a single form of a single test — and since it applies specifically to that form, variation among forms doesn't affect it. It is affected, however, by most of the other usual sources of error variance that affect split-half reliability coefficients. Among these sources are:

a. Tendency to guess when one encounters an item to which one doesn't know the answer.

b. Clerical errors in marking the answer sheet.

c. Momentary distractions.

d. Momentary fluctuations in test-taking set, motivation, attitude, tendency to guess, clerical accuracy, etc.

2. Assumptions underlying coefficient of internal precision.

   a. The basic conditions required for use of the split-half reliability procedure are also required for the coefficient of internal precision. These conditions are:

      1) That the test is unspeeded.
      2) That the items are all experimentally independent.

   b. In addition to these two requirements, there is another basic assumption of coefficients of internal precision, and this one doesn't necessarily apply to ordinary split-half reliability coefficients. It is the assumption that all the kinds of unreliability that affect coefficients of internal precision are manifested in behavior that in effect operates exactly like guessing. In other words, to the extent that a situation producing unreliable variance exists, this unreliable variance is produced in exactly the same way as if the examinee selected his response to an item absolutely at random. Thus error variance is assumed to result either from actual guessing behavior or from quasi-guessing behavior.

   c. Two assumptions that underlie the conventional correction-for-chance scoring formula are also basic to the coefficient of internal precision:

      1) The probability of getting an item right by

guessing or in any quasi-guessing circumstances is assumed to be the reciprocal of the number of options per item.

2) If an item is answered incorrectly, the assumption is that it was answered on the basis of lack of knowledge — i.e., guessing -- rather than definite misinformation.

d. Since error variance is assumed to result from either guessing behavior or quasi-guessing behavior, the error component of test scores is assumed to have, at least approximately, a binomial distribution.

e. There is still another assumption implicit in the derivation of the formula for coefficient of internal precision — and this last assumption is probably the one that seems most unusual. The formula requires the assumption that each test item is of good quality. This refers to the individual item, standing alone; there is no reference here to sampling of areas or topics to be tested. This a priori assumption that the items are structurally sound may seem to make the coefficient-of-internal-precision approach an impractical one, but it really doesn't. Rather it makes the use of certain tests an impracticable approach -- and places the responsibility of deciding whether a test is a good one right where it belongs — squarely on the researcher's shoulders. If the assumption that the items have been constructed with reasonable skill and care doesn't seem plausible after careful inspection of the test, the researcher certainly ought to think twice before using that test.

The kinds of items that make it impossible to compute coefficients of internal consistency accurately are mostly those on which the probability of guessing correctly in the absence of any knowledge at all on the point tested is much greater than is assumed in the correction-for-chance formula. A simple instance is the well-known horrible example where the stem of the item ends with the word "an" but only one of the options starts with a vowel. Some items, equally unsatisfactory, are defective in a diametrically opposite way. These are the items where even if one knows the answer none of the options states it clearly, so that the probability of answering correctly is substantially lower than 1 even when one knows the answer.

Both these categories of items -- those where the probability of guessing correctly is greater than it should be and those where the probability of getting the right answer is less than it should be -- not only make the correction-for-chance formula inaccurate but make the test itself undesirable to use.

There are other situations and other categories of items where the test is still a satisfactory one even though the conventional correction-for-chance formula may be somewhat less than completely accurate in yielding an estimate of the number of items to which the examinee knows the answer. Among these situations are:

1) Situations where the examinee has partial knowledge and can eliminate some options before guessing among the remaining ones. This is not an undesirable situation; it gives the examinee partial credit for partial knowledge.

2) Tests in which the distractors have been constructed with such diabolical cleverness that they are more attractive than the right answer when the examinee has no knowledge at all about the item. Even though such items impair the accuracy of the correction-for-chance formula they are highly desirable, because they tend to raise the reliability of the test.

3) Situations in which the examinee neglects to answer items to which he knows the answer. This somewhat aberrant behavior, whatever causes it, is unfortunate -- particularly from the viewpoint of the examinee -- because it makes the corrected score an underestimate of the amount known -- but it doesn't invalidate the coefficient of internal precision at all and is not therefore grounds for rejecting a test. As a matter of fact it doesn't even invalidate the correction-for-chance score formula, if we redefine the corrected score to mean not merely how many items the examinee knows the answer to, but rather how many he knows and knows that he knows.

In other words the corrected score may then be regarded as an estimate of the extent of the "solid core" of the examinee's knowledge.

The implications of the conventional correction-for-chance procedure are stressed here because they are basic in the derivation of the formulas for coefficient of internal precision.

To recapitulate, then, the main assumptions underlying the coefficient of internal precision are as follows:

a. The test must be unspeeded.

b. The items must be experimentally independent.

c. The error component of the test score (i.e. the unreliable component) is assumed to be due to answering items on a guessing or quasi-guessing basis and therefore to be distributed binomially.

d. Wrong answers are due to guessing and therefore contribute to the error component of the test score.

e. The individual test items should be of reasonably good quality.

C. The logic of the derivation.

Formulas for the coefficient of internal precision are shown in the Appendix, on page C-20. The derivation of these formulas involves a lot of algebra but it is fairly straightforward, involving only a very few assumptions, none of which are at all difficult to accept. There are only two points that present any problems computationally. The first of these problems is that the formulas for the coefficient of internal precision of a rights score and of any kind of formula score except that representing the conventional correction for chance, require a value of γ_{AA}, which cannot be estimated directly. Upper and lower-bound limits of γ_{AA} can be used, however, to get upper and lower-bound estimates of γ_{RR}. Fortunately the upper and lower bounds of γ_{RR} turn out to be very close together so that it doesn't really matter which is used.

The second problem in computing coefficients of internal precision is that the formulas all involve a term called $\sigma_{K_{1\infty}}$ and it is the estimation of this value that requires all of the assumptions already mentioned.

$K_{i\infty}$ is the number of items to which the examinee *knows* the correct answer, in the specific item set constituting the test. This value is defined by formula 4 in the Appendix on page C-18. Determining the standard deviation of $K_{i\infty}$ involves matrix algebra, which is summarized in the Appendix on page C-19, where the three relevant matrices are defined, and on the bottom of page C-20, where formula 12, the basic matrix formula showing the relationship among the three matrices, is presented.

The distribution of $K_{i\infty}$ is synthesized from matrix $F_{n'}$, by adding terms of the matrix to get the frequencies corresponding to various $K_{i\infty}$ values. One complication in the matrix manipulation is that sometimes, because of sampling errors which result in slight departures of actual data from the theoretical binomial distributions represented by matrix P_b, matrix $F_{n'}$ turns out to have some negative elements. This is unacceptable since the elements of $F_{n'}$ are frequencies. The procedure for getting around this difficulty is to find a new probability matrix P_w which is slightly different from P_b, the binomial matrix, and presumably represents what might be the actual distribution complete with sampling errors, instead of a theoretical binomial distribution. Using P_w instead of P_b gives a new $F_{n'}$ matrix which is free of negative terms.

D. Applicability of formulas involving reliability coefficients.

One convenient feature is that most of the usual formulas involving reliability coefficients remain sound when coefficients of internal precision are substituted, provided the circumstances are such that use of coefficients of internal precision is more appropriate than use of ordinary reliability coefficients. Examples of such formulas, where γ_{ii} can be used instead of r_{ii} are the formula for correcting for attenuation and the formula for correcting for range.

But though the coefficient of internal precision is basically a variant of the reliability coefficient it probably should not be regarded as actually being a kind of reliability coefficient. Coefficients of internal precision and ordinary reliability coefficients should be used in different situations. They are not interchangeable. In situations where one is suitable the other generally is not. For instance under no circumstances is the coefficient of internal precision a substitute for a conventional reliability coefficient in evaluating the quality of a test..

Rather, it is a special-purpose statistic, designed especially for use in evaluating the degree of relationship between the reliable components of variables that may be inferred from certain correlation coefficients.

So much for the theory. Now let's look at some empirical data.

E. Empirical data

Coefficients of internal precision have been computed on Grade 12 data for seven tests from the Project TALENT battery. The chief purpose of Table 1 is to demonstrate that the tests are essentially unspeeded and that the numbers of cases used are large enough that one can have some confidence in the results.

Table 2 shows lower and upper-bound estimates of γ_{RR} (in columns 11 and 12 respectively) and estimates of γ_{KK} in column 13. (K is the score corrected for chance. Since γ_{AA} doesn't enter into the formula for γ_{KK}, γ_{KK} can be computed directly without obtaining upper and lower-bound estimates.) The lower-bound estimates for γ_{AA} (shown in column 10) are so low that it would hardly have mattered if zeros had been used instead. For the upper-bound estimates of γ_{AA}, 1's were used. Nevertheless, the lower and upper-bound estimates of γ_{RR} hardly differ at all. Most of the differences are restricted to the third decimal place.

Table 3 shows γ_{RR} corrected for range, to give values for grade 9 and grade12 that can be applied to make a correction for attenuation, on the correlation between scores obtained in grade 9 and scores obtained by the same students three years later, in grade 12.

The correction for attenuation is shown in Table 4.

The uncorrected correlation between grades 9 and 12 is shown in Table 4, column 1. The correlations properly corrected for attenuation, using the coefficients of internal precision, are shown in column 7. They look quite reasonable. In column 6, on the other hand, we see what the correlations look like when they are improperly corrected for attenuation, using ordinary reliability coefficients for the purpose instead of coefficients of internal precision. Several of these "corrected" coefficients are over 1.

What actual effect would use of coefficients of internal precision have on heritability ratios? Let's look at R-102, the vocabulary test, on Table 4. The grade 9 reliability coefficient for girls is about .75 while the corresponding coefficient of internal precision, γ_9, is .90. This means that if the heritability ratio obtained by using the reliability coefficient were .84, the more accurate value obtained by using instead the coefficient of internal precision would be only .72. For R-109, the Scientific Attitude scale, the differences would be more extreme. A heritability ratio of .80 obtained using reliability coefficients would drop to about .50 if coefficients of internal precision were used instead. Changes of this magnitude are hardly trivial.

F. Conclusion

To summarize, in correcting correlation coefficients for attenuation and in kindred procedures to evaluate correlation, the coefficient of internal precision should usually be used instead of a conventional reliability coefficient when the two variables correlated are scores obtained on exactly the same set of test items. Two ordinary situations in which this occurs quite frequently are:

1. Heritability studies in which the correlation between twins' scores on the same test is obtained.

2. Longitudinal studies in which the same test is administered to the same people at two different times.

Since the coefficient of internal precision is systematically higher than a conventional reliability coefficient, use of the latter where the former is more appropriate makes the correlation corrected for attenuation appear to be higher than it really is. One important implications of this is that use of the coefficient of internal precision in obtaining heritability coefficients should make it apparent that the proportion of variance subject to the effects of environment is considerably higher than it has sometimes appeared to be, and the genetically determined variance is accordingly lower.

In addition to its specific implications for heritability research, the coefficient of internal precision has at least two rather general implications for psychometrics.

First, there is the assumption that the test being used is a well-constructed one. On the whole the necessity of being able to make and believe this assumption should have a rather salutary effect. If the researcher has reason to suspect that the assumption isn't true for the test he is using perhaps he will give a little more thought to the question of why he is using the test at all - and will decide to replace it with a better one. Tests that have low reliability merely because they are short are highly tolerable in many kinds of educational research, but tests that have low reliability because they are basically bad tests are unlikely to serve a useful function in any context. In most areas a knowledgeable person should be able to spot such tests by inspection.

The second general implication is that when one uses reliability coefficients it is important to know exactly what one is doing and why one is doing it. The differences in results obtained when coefficients of internal precision are used instead of ordinary reliability coefficients underlines the undesirability of ever using reliability coefficients, or numbers labeled "reliability coefficients", blindly instead of considering carefully how they were obtained, what kinds of "error variance" are included in them and what kinds are excluded, and how appropriate these inclusions and exclusions are for the specific use to which the reliability coefficient is to be applied.

APPENDIX TO

"The Coefficient of Internal Precision:
A New Psychometric Statistic for Heritability Research"

NOTATION

L = length of test (= number of test items)

n = no. of options per test item
(n = 5 for all tests used in these tables)

N = no. of cases

Kinds of scores.

(1) Scores that can actually be obtained for each individual

W = no. of wrong responses

R = raw score, not corrected for chance
(R = no. of right responses)

F = formula score, where the formula is a linear function of R and W

$$F = R + cW \tag{1}$$

where c is a constant

K = score corrected for chance by the conventional formula

$$K = R - \frac{W}{n-1} \tag{2}$$

(K is a special case of F)

A = no. of items attempted

$$A = R + W \tag{3}$$

(2) Hypothetical scores, not actually obtainable for the individual

n' = no. of items on which individual attempts to guess the answer -- i.e., number of items on which he selects an answer even though he doesn't know the correct answer.

K_∞ = mathematical expectation of K for the individual.

K_∞ = number of items to which the individual knows the correct answer.

Dual subscripts on scores.

First subscript: represents test form or set of items.

Second subscript: Represents occasion on which test is administered.

Thus A_{ij} represents no. of items attempted on item set i, on j^{th} administration of the test.

and

$A_{i\infty}$ represents hypothetical mean of no. of items attempted on a hypothetical infinite number of test administrations.

$A_{i\infty}$ = true A-score on item set i.

$K_{i\infty}$ = no. of items in item set i, to which the individual knows the correct answer.

$$K_{i\infty} = A - n' \tag{4}$$

Kinds of coefficients.

r_{XY} = actual correlation coefficient between variables X and Y.

r_{AR} = actual correlation between A-score and R-score based on the same administration of the same form of the test.

$$r_{AR} = r_{A_{11}R_{11}}$$

r_{XX} = reliability coefficient of variable X.

γ_{XY} = gamma index between variables X and Y, which are hypothesized experimentally independent variables based on the same test items; in other words: $\gamma_{XY} = r_{X_{11}Y_{12}}$

γ_{XY} = hypothetical correlation coefficient representing the correlation between measures on variable X obtained in an actual administration and measures on variable Y that might have been obtained in a different administration, hypothesized to be experimentally independent of the first one; or vice versa. The psychometric instrument (i.e., set of test items) is hypothesized to be the same for the two variables correlated.

γ_{AR} = hypothesized correlation coefficient between R-score (no. of items right) on one administration of a test and A-score (no. of items attempted) on a different administration of the same test. $\gamma_{AR} = r_{A_{11}R_{12}}$

γ_{XX} = coefficient of internal precision (gamma index) on variable X. $\gamma_{XX} = r_{X_{11}X_{12}}$

γ' = lower-bound estimate

γ'' = upper-bound estimate

Matrices

P_b = square matrix of order L+1, in which the columns, from left to right, correspond to values of n' from 0 to L, and the rows, from top to bottom, correspond to values of W from 0 to L. The elements (p) of the matrix are the corresponding probabilites in a binomial distribution on the basis of the conventional assumption about the probability of chance success.

$$P_{wn'} = \binom{n'}{w} \left(\frac{n-1}{n}\right)^{w} \left(\frac{1}{n}\right)^{n'-w} \tag{5}$$

F_w = square matrix of order L+1, in which the columns, from left to right, correspond to values of A from 0 to L, and the rows, from top to bottom, correspond to values of W from 0 to L. The elements of the matrix are empirical frequencies for the bivariate distribution of A and W.

$F_{n'}$ = square matrix of order L+1, in which the columns, from left to right, correspond to values of A from 0 to L, and the rows, from top to bottom, correspond to values of n' from 0 to L. The elements of the matrix are hypothetical frequencies for the bivariate distribution of A and n'.

FORMULAS

$$\gamma_{AA} \geq \gamma'_{AA} = \left[\frac{n r_{AR}\sigma_R - \sigma_A}{(n-1)\,\sigma_{K_{i\infty}}}\right]^2 \qquad (6)$$

$$\gamma_{RR} = \frac{1}{\sigma_R{}^2}\left[\left(\frac{n-1}{n}\right)^2 \sigma^2_{K_{i\infty}} + \frac{2}{n}\, r_{AR}\sigma_A\sigma_R + \frac{1}{n^2}\,\gamma_{AA}\sigma^2_A - \frac{2\sigma^2_A}{n^2}\right] \qquad (7)$$

$$\gamma_{KK} = \frac{\sigma^2_{K_{i\infty}}}{\sigma_K{}^2} \qquad (8)$$

$$\gamma_{FF} = \frac{1}{\sigma^2_F}\left[\frac{(1-c)^2(n-1)^2}{n^2}\,\sigma^2_{K_{i\infty}} + \left(\frac{nc-c+1}{n^2}\right)^2 \gamma_{AA}\sigma^2_A + \frac{2(1-c)(nc-c+1)}{n}\left(r_{AR}\sigma_A\sigma_R - \frac{\sigma^2_A}{n}\right)\right] \qquad ($$

$$\gamma_{RR} \leq \gamma^{11}_{RR} = \frac{1}{\sigma^2_R}\left[\left(\frac{n-1}{n}\right)^2 \sigma^2_{K_{i\infty}} + \frac{2}{n}\, r_{AR}\sigma_A\sigma_R - \frac{\sigma^2_A}{n^2}\right] \qquad (10)$$

$$\gamma_{RR} \geq \gamma^{1}_{RR} = \frac{1}{\sigma^2_R}\left[\left(\frac{n-1}{n}\right)^2 \sigma^2_{K_{i\infty}} + \left(\frac{n r_{AR}\sigma_A\sigma_R - \sigma^2_A}{n(n-1)\sigma_{K_{i\infty}}}\right)^2 + \frac{2}{n}\, r_{AR}\sigma_A\sigma_R - \frac{2\sigma^2_A}{n^2}\right] \qquad (11)$$

Determining $\sigma_{K_{i\infty}}$

Matrix formula: $F_{n'} = P_b^{-1}\ F_w$ (12)

Distribution of $K_{i\infty}$ is obtained by applying formula 4 to Matrix $F_{n'}$ (as obtained from formula 12).

Table 1. Basic descriptive data on the variables and the sample.

Variable #	Variable	No. of Items	No. of cases[*]		% finishing[**] (1963 retest data Grade 12)	
			M	F	M	F
102	Vocabulary I	21	1930	2000	95.8	98.5
106	Math Info.	23	1930	2000	95.1	98.3
109	Scientific Attitude	10	1930	2000	99.7	99.9
112	Mech. Info.	19	1930	2000	96.1	98.5
113	Farming Info.	12	1930	2000	97.8	99.3
231	Spelling	16	1835	1963	95.7+	97.5+
290	Abstract Reasoning	15	3670	3861	97.2	97.0

---

[*] The cases used in this study were cases from the TALENT retest study, which is reported on in:

> Shaycoft, Marion F. The High School Years: Growth in Cognitive Skills, 1967, (Interim report 3 to the U. S. Office of Education, Cooperative Research Project No. 3051.) Pittsburgh: American Institutes for Research and Univer. of Pittsburgh, Project TALENT Office.

[**] Ibid, Tables 4-1 and 4-4. These data for percent finishing are based on largely but not exactly the same groups as the ones described in the two "No.of cases" columns above.

Table 2. Gamma indexes and other relevant statistics based on grade 12 data

Sex	Variable #	(1) r_{AR}	(2) σ_R	(3) σ_A	(4) σ_K	(5) $\sigma_{K_{1\infty}}$	(6) $\bar{R}$	(7) $\bar{A}$	(8) $\bar{K}$	(9) $\bar{K}_{1\infty}$	(10) γ'_{AA}	(11) γ'_{RR}	(12) γ''_{RR}	(13) γ_{KK}	(14) N
M	102	.323	3.64	1.51	4.58	4.23	14.53	20.36	13.07	13.07	.067	.905	.911	.852	1930
	106	.306	6.36	2.19	7.97	7.59	12.11	22.20	9.59	9.59	.062	.945	.950	.907	1930
	109	.245	1.86	.56	2.34	2.08	6.78	9.86	6.01	6.01	.043	.818	.822	.790	1930
	112	.394	3.05	1.29	3.82	3.48	13.61	18.55	12.38	12.38	.115	.887	.893	.829	1930
	113	.317	2.11	.84	2.66	2.38	8.50	11.73	7.69	7.70	.069	.849	.855	.800	1930
	231	.275	3.12	1.31	3.94	3.55	9.51	15.83	7.93	7.94	.044	.863	.870	.811	1835
	290	.112	2.72	.43	3.40	3.16	10.17	14.93	8.98	8.98	.007	.870	.871	.862	3670
F	102	.273	4.04	1.47	5.08	4.73	13.35	20.37	11.60	11.60	.045	.908	.913	.867	2000
	106	.263	5.90	2.68	7.50	6.94	9.07	21.80	5.89	5.89	.034	.918	.926	.857	2000
	109	.224	1.83	.51	2.29	2.04	6.79	9.87	6.02	6.02	.036	.813	.815	.789	2000
	112	.303	2.91	2.22	3.88	3.07	8.85	17.97	6.57	6.57	.032	.761	.784	.627	2000
	113	.252	2.29	.84	2.89	2.60	7.87	11.72	6.91	6.91	.039	.848	.853	.808	2000
	231	.324	2.68	1.15	3.38	3.03	11.13	15.86	9.95	9.95	.070	.861	.868	.805	1963
	290	.125	2.78	.58	3.49	3.22	9.78	14.91	8.50	8.50	.008	.865	.867	.852	3861

C-22

γ'_{AA} is obtained from formula 6. This is a weak lower-bound estimate.

γ'_{RR} is a lower-bound estimate obtained by using formula 6 for γ'_{AA}. (See formula 11.)

γ''_{RR} is an upper-bound estimate obtained by using $\gamma''_{AA} = 1$. (See formula 10.)

Table 3. Correction of γ'_{RR} for range, to determine γ_9 and γ_{12} corresponding to $r_{9,12}$ in Table 4.

Sex	Variable #	(1) γ'_{RR}	(2) σ_R	(3) σ_{R_9}	(4) γ_9	(5) $\sigma_{R_{12}}$	(6) γ_{12}
M	R-102	.905	3.64	3.80	.913	3.71	.909
	R-106	.945	6.36	4.03	.863	6.35	.945
	R-109	.818	1.86	1.99	.841	1.91	.827
	R-112	.887	3.05	3.28	.902	3.15	.894
	R-113	.849	2.11	2.28	.870	2.12	.850
	R-231	.863	3.12	2.87	.838	3.12	.863
	R-290	.870	2.72	3.00	.893	2.84	.881
F	R-102	.908	4.04	3.86	.900	4.09	.911
	R-106	.918	5.90	3.61	.782	5.88	.918
	R-109	.813	1.83	1.86	.819	1.83	.813
	R-112	.761	2.91	2.72	.727	3.02	.779
	R-113	.848	2.29	2.51	.873	2.30	.849
	R-231	.861	2.68	2.74	.867	2.66	.859
	R-290	.865	2.78	2.96	.881	2.95	.880

Columns 1 and 2 of this table are columns 11 and 2 respectively of Table 2.

The standard deviations in columns 3 and 5 are from columns 9-10 and 11-12 respectively of Table 4-9 of (Shaycoft, op. cit.)

Columns 4 and 6 represent corrections of γ'_{RR} (col. 1) for range, to correspond to the standard deviations of columns 3 and 5 respectively.

Table 4. Correction of $r_{9,12}$ for attenuation

Sex	Variable #	(1) Inter-grade correlation $r_{9,12}$	(2) Reliability coefficients r_9	(3) r_{12}	(4) Coefficients of int. precision γ_9	(5) γ_{12}	(6) $\frac{r_{9,12}}{\sqrt{r_9 r_{12}}}$	(7) $r_{9_{i\infty}12_{i\infty}}$
M	R-102	.741	.786	.776	.913	.909	.95	.814
	R-106	.716	.749	.899	.863	.945	.87	.793
	R-109	.552	.594	.559	.841	.827	.96	.662
	R-112	.648	.728	.706	.902	.894	.90	.722
	F-113	.628	.659	.602	.870	.850	1.00	.730
	R-231	.640	.672	.721	.838	.863	.92	.753
	R-290	.568	.746	.717	.893	.881	.78	.640
F	R-102	.764	.748	.776	.900	.911	1.00	.844
	R-106	.686	.710	.891	.782	.918	.86	.810
	R-109	.519	.520	.504	.819	.813	1.01	.636
	R-112	.575	.504	.595	.727	.779	1.05	.764
	R-113	.696	.686	.626	.873	.849	1.06	.808
	R-231	.590	.659	.638	.867	.859	.91	.684
	R-290	.601	.734	.731	.881	.880	.82	.682

Column 1 is from columns 1-2 of Table 4-9 of (Shaycoft, op. cit.)

Columns 2 and 3 are from columns 1-2 and 3-4 of Table 4-8 (ibid.)

Columns 4 and 5 are from Table 3

Column 6 represents "correction of $r_{9,12}$ for attenuation" by the conventional formula, which is inappropriate in this situation. It results in overestimates

Column 7 represents correction of $r_{9,12}$ for attenuation by a more appropriate formula, which uses coefficients of internal precision instead of ordinary reliability coefficients. The formula is:

$$r_{9_{i\infty}12_{i\infty}} = \frac{r_{9,12}}{\sqrt{\gamma_9 \gamma_{12}}}$$

Appendix D1

MUST ALL MEMBERS OF A PROBABILITY SAMPLE BE RANDOMLY SELECTED? *

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

A. PROBLEM

Sometimes in a research study that calls for selection of a probability sample of schools (or educational programs, or individuals, or any other kind of entities to be selected from a known finite popolation), for one reason or another additional schools (or programs or individuals) have to be added to the study. This may be because in addition to the phase of the study in which characteristics of the probability sample are to be investigated there are some schools (or programs or individuals) of particular interest that the researcher wishes to pay particular attention to -- perhaps to handle as case studies in a separate phase of the investigation. Suppose, for instance, that the government wished a study to be made of education programs funded under ESEA Title XLIII**, which has as its purpose the development of educational programs to help children handicapped by sinistralism to overcome their "disability". Suppose further that it was desired that this study be conducted in two phases -- the first being a study of a probability sample of all Title XLIII programs, and the second consisting of intensive studies of a small number of such programs designated for inclusion because on the basis of preliminary evidence they seem particularly promising. Another situation that sometimes occurs in which schools have to be added to the study after the probability sample has been drawn is when certain schools drawn in the probability sample selection agree to participate only on condition that certain other schools also be included.***

---

* This is a slightly revised version of a paper that was presented at the American Psychological Association meeting in Washington, D.C., on 7 September 1976.

** This is a purely hypothetical situation; there is no Title XLIII in the ESEA -- and certainly no program for sinistralism!

*** For instance in Project TALENT something like that sometimes occurred when one of the two secondary schools in a particular district was selected in a stratified random sampling procedure. The school superintendent was interested in having the selected school participate, because he (or she) wanted to have the report of results that Project TALENT would provide for the participants -- but he wanted the results for all the secondary school students in his (or her) jurisdiction, not just ones in the school that happened to have been selected. Therefore he would make testing of the other school as well a condition of participation.

Under such circumstances careful researchers who have found it necessary to include schools not selected by the random-number procedure that was used to selected the regular probability sample have customarily excluded the extra schools from that probability sample. Data from the extra schools, while available for use in "case studies" or other types of data analysis in which a probability sample wasn't required, were excluded from any analyses in which the kinds of inferences that required a probability sample were to be drawn.

But is this really the best way to proceed? If all the expense of including the "extra" schools is incurred, isn't it extravagantly wasteful to then ignore the data from them (except for certain very limited uses)? Is there any way these schools, not selected by random means as the probability sample schools were, can be incorporated in the probability sample without doing violence to the laws of probability sampling?

B. GENERAL NATURE OF THE SOLUTION

The answer to the question posed above is that it is both possible and desirable to incorporate non-randomly selected schools into a probability sample selected by traditional random means from a finite population. To do this requires that finite-population statistics and theory be used since it appears that infinite-population statistics and theory cannot be properly adapted to cope with the situation.

The solution is applicable regardless of whether the initial randomly selected probability sample was selected by simple random sampling or stratified random sampling. In the case of the former, however, the procedure in effect converts the sample to a stratified one, since the arbitrarily selected schools (in other words those not selected by random means) have to be treated as a separate stratum.

C. SPECIFIC PROCEDURE

If simple random sampling has been used, a separate stratum is created for the schools to be arbitrarily included. If stratified random sampling has been used the arbitrarily selected schools are pulled out of the stratum population in which they originally were, and transferred to one or more newly defined strata. Actually one stratum is always adequate for the purpose. There is no objection to using more than one (for the arbitrarily included schools) if it is convenient to do so, but there is no theoretical reason requiring it. Since a separate sampling ratio may be set for each stratum, it is a simple matter to set a sampling ratio of 1.00 for any stratum composed of arbitrariliy selected schools. And it is not only simple to do this, but it is also entirely legitimate, since the only restriction on sampling ratios to be used in drawing a bona fide probability sample is that each sampling ratio must be greater than zero. There is no rule that each sampling ratio must also be less than 1.00.

Of course appropriate weights must be applied to each sample member in computing statistics used as estimates of corresponding population parameters. The weight equals the reciprocal of the sampling ratio, or in the case of multistage sampling it equals the reciprocal of the product of the sampling ratios used at the various stages. This means, of course, that since arbitrarily designated sample members are conceptualized to have been selected with a sampling ratio of 1 at every stage, they will have a weight on only 1 -- the smallest weight possible. Thus, although these arbitrary sample members will affect the sample statistics they will generally have considerably less effect than randomly selected population members.

D. NOTATION USED IN ALGEBRAIC DEVELOPMENT

g = number of "groups" or "strata" or "cells" from which sampling units are selected randomly (a separate sample from each group).

j = number of "groups" or "strata" or "cells" from which sampling units are selected judgmentally (i.e., arbitrarily).

Note that for purposes of sample selection there is no reason why j ever need be greater than 1. All sampling units that are to be selected judgmentally rather than randomly can be arbitrarily combined into a single stratum or group, from which 100 percent of the sampling units are to be selected.

N_i = number of sampling units in stratum i population before identification of the judgmentally selected sampling units.

N = total number of sampling units in population

$$N = \sum_{i=1}^{g} N_i \tag{1}$$

n_i = number of sampling units already selected from stratification cell i before identification of the judgmentally selected sampling units.

$(i = 1,2,3,\ldots g)$

Note that if $N_i > 0$, n_i must be greater than 0.

$$n = \sum_{i=1}^{g} n_i \tag{2}$$

c_i = number of judgmentally selected sampling units that were originally in stratification cell i

J = total number of sampling units in the population (and sample) in the j stratification cells consisting of judgmentally selected units.

$$J = \sum_{i=1}^{g} c_i \tag{3}$$

d_i = number of judgmentally selected sampling units that not only were originally in stratification cell i but were among the n_i sampling units originally selected for the sample.

r_i = number of sampling units from among the d_i defined above, that are replaced randomly after being transferred to one of the strata for judgmentally selected sampling units.

(r_i can have any value between 0 and d_i.)

N_i' = revised number of sampling units in stratification cell i population after the judgmentally selected sampling units have been transferred to a new stratum.

n_i' = revised number of sampling units in stratification cell i sample.

Note that if $N_i' > 0$, n_i' must be greater than 0.

p_i = original sampling ratio for cell i

$$p_i = \frac{n_i}{N_i} \tag{4}$$

p_i' = revised sampling ratio for cell i

$$p_i' = \frac{n_i'}{N_i'} \tag{5}$$

w_i = weight for each sample member in cell i

$\bar{X}_i$ = sample mean for stratum i, based on n_i' cases, and

$\bar{X}'$ = mean of the J nonrandomly selected cases in the sample.

s_i = sample standard deviation for stratum i

$$s_i = \sqrt{\frac{\sum_{k=1}^{n_i'} (X_{ik} - \bar{X}_i)^2}{n_i'}} \qquad (6)$$

where X_{ik} = k^{th} sample member of the n_i' from stratum i.

$\sigma_{\bar{x}}$ = standard error of overall mean.

$\sigma_{\bar{x}_i}$ = standard error of stratum i mean based on the n_i' cases in the stratum after the non-randomly selected ones have been eliminated from it (and possibly replaced).

Special notation for stratified random cluster sampling

n_{ik} = number of sample members (e.g., students) in randomly selected sampling unit k (e.g., school k) in stratum i.

n_{ih} = number of sample members (e.g., students) in judgmentally selected sampling unit h (e.g., school h) originally in stratum i.

N" = estimated total number of students in population.

$\bar{X}_{ik}$ = mean value of variable X based on the n_{ik} sample members (e.g., students) in sampling unit k in stratum i.

$\bar{X}_{ih}$ = mean value of variable X based on the n_{ik} sample members (e.g. students) in sampling unit h originally in stratum i.

E. PROCEDURES FOR SELECTING SAMPLE AND DETERMINING WEIGHTS

The researcher can proceed in either of two ways, designated "replacement procedure" and "nonreplacement procedure" respectively, or by a third procedure, which is a combination of the first two.

Replacement procedure

a. Conceptually remove the c_i judgmentally selected sampling units from the cells containing them.

b. Randomly draw d_i additional sampling units from each such cell i to replace the d_i previously drawn sampling units that have been removed from the cell.

c. The sampling ratio for each of the n_i randomly selected sampling units representing cell i becomes:

$$p'_i = \frac{n_i}{N_i - c_i} \qquad (7a)$$

Nonreplacement procedure

a. Conceptually remove the c_i judgmentally selected sampling units from the cells containing them. Do not replace any of them.

b. The sampling ratio for each of the $n_i - d_i$ randomly selected sampling units remaining to represent cell i becomes:

$$p'_i = \frac{n_i - d_i}{N_i - c_i} \qquad (7b)$$

Generalized formula, covering both procedures and a possible cross between them

A cross between the two procedures might occur if some but not all of the d_i judgmentally selected sampling units are replaced.

Let r_i = number of sampling units replaced in cell i.

(r_i can have any value between 0 and d_i.)

For the replacement procedure, $r_i = d_i$.

For the nonreplacement procedure, $r_i = 0$.

For a cross between the two procedures, r_i is somewhere between those two values; in other words, $d_i > r_i > 0$.

The generalized formula covering all three possibilities is:

$$p_i' = \frac{n_i - d_i + r_i}{N_i - c_i} \qquad (7c)$$

It should be noted that if $c_i = N_i$ stratum i in effect disappears from the population. This creates no problems. But if $d_i = n_i < N_i$ the nonreplacement procedure must not be used for that stratum; otherwise the stratum would be represented in the population but not in the sample. Therefore, if $d_i = n_i < N_i$ it is essential that at least some of the d_i sampling units withdrawn from the part of the sample representing stratum i be replaced; in other words that:

$$r_i > 0$$

It should also be noted that for any stratum for which $N_i' > 1$, it is essential that $n_i' > 1$. Otherwise the standard error for the mean would be indeterminate (see formula 21).

Regardless of which of the three sample selection procedures is used, the formula for the weight to be applied to each sample member is:

$$w_i = \frac{1}{p_i'} \qquad (8)$$

F. ESTIMATING THE POPULATION MEAN

1. For stratified random sample (from finite population)

Using the generalized procedure as described in Section A2c above, the population size for stratum i (i = 1,2,3,...g) is:

$$N_i' = N_i - c_i \qquad (9)$$

and the sample size for stratum i (i = 1,2,3,...g) is:

$$n_i' = n_i - d_i + r_i \qquad (10)$$

The formula for estimating the population mean is:

$$\bar{X} = \frac{\sum_{i=1}^{g} w_i \, n_i' \, \bar{X}_i + \sum_{i=1}^{g} \sum_{k=1}^{c_i} X_{ik}}{N} \qquad (11)$$

where X_{ik} = value of variable X for k^{th} nonrandomly selected sample member originally in the stratum i population.

$$\bar{X} = \frac{\sum_{i=1}^{g} N_i' \, \bar{X}_i + J \, \bar{X}'}{N} \qquad (12)$$

where N_i' and J are defined in formulas 9 and 3 respectively.

2. For stratified random cluster sample (from finite population)

This would occur when the sampling unit is not the individual case but a "cluster" of cases. If the cluster is selected for the sample, all the cases in it are elements of the sample. One example would be the use of schools as "clusters". The school would be the basic sampling unit, and all the students in it would be members (elements) of the sample.

The weights used are the same weights (i.e., those given by Formula 8) for a cluster sample as for a one-stage non-cluster sample.

Thus the estimated number of sample members is:

$$N'' = \sum_{i=1}^{g} \sum_{k=1}^{n_i'} w_i \, n_{ik} + \sum_{i=1}^{g} \sum_{h=1}^{c_i} n_{ih} \qquad (13)$$

The estimated population mean of X, estimated on the basis of the sample members (e.g., students) is:

$$\bar{X} = \frac{\sum_{i=1}^{g} \sum_{k=1}^{n_i'} w_i \, n_{ik} \, \bar{X}_{ik} + \sum_{i=1}^{g} \sum_{h=1}^{c_i} n_{ih} \, \bar{X}_{ih}}{N''} \qquad (14)$$

G. EFFECT ON SAMPLING ERROR

As has been pointed out, inclusion of arbitrarily selected cases in the probability sample will usually not have a large effect on the statistics computed. But what effect inclusion of these cases does have will be particularly valuable because it will contribute absolutely nothing to sampling error. This assertion can be made because in the case of arbitrarily selected sample members (i.e. members selected with a sampling ratio of 1), stratum population and stratum sample coincide exactly, reducing sampling error for the stratum to 0.

The standard error of a mean based on a simple random sample from a finite population is given by formula 15.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}} \qquad (15)$$

where N = number of cases in population

n = number of cases in sample

s = sample standard deviation

$$s = \sqrt{\frac{\sum_{1}^{n} (X - \bar{X})^2}{n}} \qquad (16)$$

when formula 15 is applied to a stratum mean where sample and population coincide, $N - n = 0$, and clearly the standard error of the stratum mean reduces to 0.

Substituting the stratum values s_i, n_i', and N_i' for s, n, and N in formula 15 yields:

$$\sigma_{\bar{x}_i} = \frac{s_i}{\sqrt{n_i'-1}} \sqrt{\frac{N_i' - n_i'}{N_i'}} \qquad (17)$$

For any stratum for which sample and population coincide. $N_i - n_i = 0$, so that clearly such a stratum contributes nothing at all to the standard error of the overall mean. Thus $\sigma_{\bar{x}'} = 0$, as shown by applying formula 17 to the j strata containing the J nonrandomly selected sampling units:

$$\sigma_{\bar{x}'} = \frac{s_j}{\sqrt{J-1}} \sqrt{\frac{J - J}{J}} = 0 \qquad (18)$$

Using as a starting point the usual formula for standard deviation of a sum, we see that the standard error of a stratified random sample mean computed from formula 12 is:

$$\sigma_{\bar{x}} = \frac{1}{N} \sqrt{\sum_{i=1}^{g} \sigma^2_{N_i'\bar{x}_i} + \sigma^2_{J\bar{x}'}} \qquad (19)$$

$$= \frac{1}{N} \sqrt{\sum_{i=1}^{g} N_i'^2 \sigma^2_{\bar{x}_i} + J^2 \sigma^2_{\bar{x}'}} \qquad (20)$$

Substituting (17) and (18) in (20), the standard error of a stratified random sample mean (computed from formula 12) is:

$$\sigma_{\bar{x}} = \frac{1}{N} \sqrt{\sum_{i=1}^{g} \frac{N_i'^2}{n_i'-1} \frac{N_i'-n_i'}{N_i'} s_i^2 + 0}$$

$$= \frac{1}{N} \sqrt{\sum_{i=1}^{g} \frac{N_i'}{n_i'-1} (N_i'-n_i') s_i^2} \qquad (21)$$

It should be noted that formula 21 is not applicable to a stratified random cluster sample. The standard error of the mean of such a sample has not as yet been derived by the author.*

H. APPLICABILITY AND ADVANTAGES OF THE PROPOSED PROCEDURE

The procedure described is useful whenever three conditions are met:

1. First: that a probability sample (either simple random or stratified random) is to be drawn.
2. Second: that the population may be regarded as finite.
3. Third: that specified members of the population are to be in the sample whether or not they are selected by whatever random procedure is being used.

The procedure has three intertwined advantages:

1. First: it makes maximum use of available data.
2. Second: it results in some reduction in sampling errors. (Although this reduction, in most situations, is not likely to be large, it will nevertheless be worthwhile -- particularly in view of advantage number 3.)
3. Third: advantages 1 and 2 are achieved at virtually no cost (assuming, of course, that the situation is such that the designated population members are to be included in the group on whom data is collected whether or not they are incorporated in the probability sample).

---

* The derivation presents certain difficulties since in the case of cluster samples both the number of cases in the cluster and the means are subject to sampling errors, rather than just the means. Thus the sampling errors of products are involved.

Appendix D2

THE "CLOSED SCHOOL-CLUSTER" METHOD OF SELECTING A PROBABILITY SAMPLE*

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

A. The problem

This paper is concerned with a problem that applies to many educational research studies and that hardly ever (or possibly never) has been handled optimally. The problem is how to select a probability sample of schools in such a way that all grade levels to be included in the study are represented either by the same schools or by corresponding schools; for instance the high schools that provide the grade 10 sample should be those high schools whose students come from the elementary schools that provide the grade 3 sample. The sample to be selected should not only meet this requirement but it should do it in a relatively efficient way; in other words in such a way that the sampling errors are kept as small as practical considerations will permit. Practical considerations also dictate that the procedure should be cost-effective.

B. Applicability of the problem

The problem is an important one because the studies to which it is particularly applicable are the large-scale expensive-to-execute ones that are either ambitious longitudinal studies or cross-sectional studies covering a fairly wide range of grades or both. The kind of longitudinal studies to which the sampling problem we are here concerned with applies are those that call for following the same cohort of students over a period of several years, in the course of which two or more school levels (e.g., elementary school, junior high school, senior high school) are involved. The kind of cross-sectional studies to which the sampling problem applies are those for which the samples for all grades should be composed of either the identical schools or insofar as different levels of schools (e.g., both elementary and secondary schools) are required, the corresponding schools in the same school system. The type of design that Beaton refers to as a "pseudo longitudinal" design* would fall in this category. For such studies, as well as for true longitudinal studies, it is generally desirable that rather than selecting an entirely separate sample of schools at each school level (for instance a probability sample of elementary schools and a separately selected probability sample of secondary schools) the schools at the different levels should be selected in such a way as to "correspond." This requirement may be necessary for reasons of administrative convenience and economy--in other words fewer school systems to deal with--or it may be necessary because the research plan calls for comparison of different grade cohorts at the same time within the same (or corresponding) schools.

---

*This is a slightly revised version of a paper presented at the Annual Meeting of American Educational Research Association in Washington, 1 April 1975. The author is indebted to Dr. Charles Stegman for some helpful suggestions concerning the earlier version.

Project TALENT is an exemplar of a study that is both cross-sectional, in the manner we have been talking about, and longitudinal. It cuts across several grades (grades 9-12) and involves corresponding schools at two levels--primarily senior high schools (for grades 10-12 and for some grade 9 students) but some junior high schools (for those grade 9 students not in four-year high schools). The Project TALENT sample was not selected in precisely the manner to be proposed in this paper--a manner which utilizes a concept we are terming "closed clusters" of schools. (We hadn't developed the concept back in 1959, when selection of the sample was being planned.) As a matter of fact, although the Project TALENT sample is generally regarded as excellent not merely by those of us who were involved in its design and selection but by researchers entirely unconnected with the project, nevertheless we have been aware that like most products of human effort it is imperfect; the chief difficulty being that because of problems in matching junior high schools with senior high schools the grade 9 sample did not entirely correspond to the sample of students in grades 10, 11, and 12. Now we know a solution to the problem; if we were selecting the Project TALENT sample today, we would use the "closed-cluster" concept, and thereby would obtain not just a very representative sample of 10th, 11th, and 12th-grade students, which we also acheived in 1959, but also a grade 9 sample that would correspond perfectly to the grade 10-12 sample in terms of school composition.

---

*Beaton, A. E. Pseudo Longitudinal Model, 1974 (unpublished paper)

C. Purpose of this paper

It is the purpose of this paper to describe a sampling procedure utilizing the concept of closed clusters of schools--a concept alluded to in the preceding paragraph--and to explain how it can be applied to solve, in a cost-effective way, the sampling problem that has been described.

D. The necessity of, and difficulty in, identifying corresponding schools at different levels

A basic question that has to be answered very early in the planning of any probability sampling for use in educational research is what the primary sampling unit should be. Should it be students? classrooms? schools? school districts? school systems? counties? states? In general, the larger the primary sampling unit the easier and more economical it is to handle administrative arrangements; on the other hand the larger the primary sampling unit the larger the sampling errors. Balancing the pros and cons, in a large-scale study to take place in the schools it is generally not feasible from a practical viewpoint to use students as the primary sampling unit. Any unit smaller than the school is likely to be too costly to be practicable.

Assuming that in the case of a particular longitudinal study the exigencies of the situation are such that the basic unit must be the school (or something larger) rather than the individual student, how should the schools be selected? This problem presents unique difficulties in longitudinal studies that span a time period when some members of the sample may move from one school to another (from elementary to high school, for instance). The problem of how to handle the shift from elementary to high school--or in districts that have junior high schools, the shift from elementary to junior high school and then to senior high school--would be no problem at all if there were a one-to-one correspondence between elementary schools and high schools (and junior high schools where applicable). By one-to-one correspondence we mean that all graduates of elementary school a go to high school A (if they go to high school at all); similarly that all graduates of elementary school b go to high school B; that the graduates of elementary school c go to high school C; and conversely that high school A normally draws students from no elementary school but a; high school B from no elementary school but b, etc. These rules refer, of course, just to normal school-to-school transitions; exceptions resulting from moving out of the neighborhood or from abnormal circumstances of one kind or another that occur rarely and just affect one or two individual children don't count in this context of defining one-to-one correspondence. But unfortunately (from the viewpoint of sampling simplicity), although one-to-one correspondence is the situation in many communities there are many others where it isn't. One way of coping with this problem is by means of the "closed cluster of schools" concept.

E. The "closed cluster of schools" concept

To begin with, let's consider the situation where we are concerned only with elementary schools and senior high schools--not with junior high schools. Such a situation might be a longitudinal study whose design involved contacting the sample members initially in grade 1 and then again in grade 5 and then in grade 11, so that none of the two junior high school grades would be involved.

In terms of the relationship between elementary and high schools (what elementary schools each high school draws from and what high schools each elementary school sends its graduates to), we may classify schools into four kinds of elementary-and-high-school cluster, as follows:

Type A cluster

This occurs where there is a one-to-one correspondence (as described above) between elementary and high schools. In this case each elementary school and its corresponding high school constitute a two-school cluster.

Type B cluster

This occurs where there is an m-to-one correspondence between elementary schools and high schools. This is the situation where several elementary schools send all their graduates to one high school and that high school draws from no other elementary schools. In this case the m elementary schools and one high school constitute an m+1 school cluster.

Type C cluster

This occurs when there is a one-to-n correspondence between elementary and high schools. It occurs when one elementary school sends its graduates to several high schools and each of those high schools draws from no other elementary school. In this case the one elementary school and the n high schools constitute a 1+n school cluster.

Type D cluster

A Type D cluster is an m+n school cluster consisting of m elementary schools and the n high schools they send their graduates to, these n high schools drawing from no other elementary schools than those m. Type D clusters, by definition, consist only of schools that do not fit into any Type A, B, or C clusters. Furthermore each Type D cluster should contain the fewest possible schools that will fit the definition. In other words no Type D cluster should consist of the combination of two or more smaller Type D clusters nor should it consist of two or more smaller clusters of any other type than Type D.

As defined above, every elementary school and every high school in the United States belongs to one and only one cluster. The classification principle can of course be readily generalized to include junior high schools. Then, clearly, every elementary school, every junior high school, and every high school in the United States belongs to one and only one cluster. We are calling these groups of schools "closed clusters" because nobody enters or leaves them except through entering grade 1, graduating from high school, moving into or out of the geographic area covered by the cluster, or becoming a dropout.

This classification of schools into clusters is at the heart of the proposed solution to the problem of sampling in such a way as to solve the problems caused by irregular patterns of movement from elementary school to secondary school, and from junior high school to senior high school.

F. Using closed clusters of schools in sample selection

It is probably obvious that the proposed solution uses closed clusters of schools as the primary sampling unit. Ideally this would involve sampling from a directory that contains a list of all such clusters and identifies the schools in each. Unfortunately, however, no such directory exists.

It is necessary, therefore, to improvise a method that will permit obtaining essentially the same results in the absence of a comprehensive list of clusters. The suggested procedure may be described as a one-stage or multi-stage stage sampling procedure, the first stage of which is a modified form of stratified or simple random cluster-sampling with differential sampling ratios.

The notation used in the description that follows is summarized in Figure 1.

The steps are as follows:

1. If the sample is to be stratified select the stratification variables and define the strata. Set the sampling ratio at approximately the level that might be expected to give about the desired number of schools if the sampling ratio were to be applied to entire clusters rather than to individual schools. (If, for instance, you surmise that the average closed cluster in the segment of the population of schools you are interested in contains about five schools, you would set your initial sampling ratio at only about one-fifth of what you would normally have used to get the number of schools you would like included. If differential sampling ratios are to be used for different strata, these, too, should be set as part of this initial step.
2. Using the stratification (if any) and the sampling ratio (or ratios) decided upon and using the Office of Education source data on public schools (or any other source you may prefer), select* the initial sample of schools. These are the potential "nucleus schools." A nucleus school is the first school selected for a cluster--in other words the nucleus of the cluster.

---

*with replacement

3. For each of the potential nucleus schools selected, determine what other schools are in the same closed cluster. This will necessitate determining, by interview, questionnaire, or delving into official records, what schools it gets its students from (if it is a secondary school) and what schools it sends its graduates to (if it is an elementary or junior high school). This information will have to be secured not only from the initial school selected (the nucleus of the cluster) but also from each school from which it receives or to which it sends students, and in turn from each of those schools, and so forth until no new schools enter the picture.

4. If multi-stage sampling has been decided upon, for instance a second-stage sampling to select a probability sample of students within the sample schools and in the appropriate grades, this should be done at this point. It is often desirable to set sampling ratios which will make the weights (w_{jk}) turn out to be at least roughly equal, since for a sample of given size this reduces the sampling errors of statistics.

5. After the sample has been selected it is necessary to determine the weight to be attached to each student in the sample. Normally the weight would be the reciprocal of the product of all the sampling ratios involved (first-stage, second-stage, etc.). With the suggested closed-cluster procedure, however, it is necessary to remember that p'_{jk} is an understatement of the probability that school i will be in the sample, because it will be in the sample not only if it is the first school selected in the sample of potential nucleus schools but also if it is a subsequent selection in the nucleus school sample; and also because it will be included in the sample if any other school in its cluster is so selected.

   The relevant formulas for determing the weights are shown in Figure 1.

Hypothetical data showing how sets of schools break down into closed clusters, and how the proposed procedure works, are shown in Tables 1-4.

Table 1 shows hypothetical percentage distributions corresponding to the transition between junior high school and senior high school in a particular community. This hypothetical community has 12 junior high schools and 10 senior high schools, and it just happens to include closed clusters of all four types--Type A, Type B, Type C, and Type D--although the researchers wouldn't have to know that in order to proceed.

Table 2 shows diagrammatically how the first-stage sampling would be done, assuming it had been decided to start with a potential-nucleus-school sample consisting of three schools from the community.

Table 3 shows how the weights would be determined, assuming a second-stage sampling is also carried out.

Table 4 is not relevant to actual execution of the procedure since it shows, on the basis of the Table 1 data, the classification of all the schools into closed clusters--not just those schools in the sample. The 22 schools of the community break down into 5 closed clusters, having 2 to 8 schools each.

G. Importance of the procedure

It is believed that the proposed procedure provides a definitive solution to the knotty problem of selecting a probability sample of schools that include all corresponding schools at different levels. This will be particularly helpful in longitudinal studies requiring a probability sample; in quasi-longitudinal studies; and in some large-scale cross-sectional studies (those that require, for administrative or research reasons, that corresponding schools at different levels be included).

FIGURE 1. Notation and formulas

NOTATION

N' = no. of schools in pool from which potential nucleus schools are selected.

m = no. of strata

N'_{hk} = no. of schools in stratum h (which contains school k) in pool from which, in the case of stratified random sampling, potential nucleus schools are selected

n' = no. of potential nucleus schools selected (including those that may be rejected because another school in the same cluster has already been selected as the nucleus school)

n'_h = no. of potential nucleus schools selected (in the case of stratified random sampling) from stratum h

P'_{jk} = probability that school k, which happens to be in cluster j, will be the first school selected in the first-stage sampling to select potential nucleus schools. (In other words p'_{jk} = probability of selection of school k on a single draw.)

P'_{hjk} = probability that school k, which happens to be in stratum h in cluster j, will be the first school selected from stratum h in the first-stage sampling to select potential nucleus schools. (In other words, p'_{jk} = probability of selection of school k on a single draw.)

P_{hj} = probability that at least one school in stratum h in cluster j will be the first school selected from stratum h in the first-stage sampling to select potential nucleus schools.

P_{1jk} = first-stage sampling ratio for selecting school k in cluster j as a nucleus school

n_j = no. of schools in cluster j

n_{hj} = no of schools in stratum h in cluster j

(Note that not all schools in the cluster are necessarily in the same stratum.)

P_{ijk} = i^{th}-stage sampling ratio ($i > 1$) applied in school k in cluster j to select a probability sample of classrooms or students (or some other entities) within school. (This is applicable only if multi-stage sampling is being done.)

P'_{1j} = probability that schools in cluster j will be in the sample (on the basis of first-stage sampling)

w_{jk} = weight for each case in school k in cluster j

FORMULAS*

$$P'_{hjk} = \frac{1}{N'_{hk}} \quad \text{(assuming all schools in the stratum have equal a priori probabilities)} \tag{1}$$

$$P_{hj} = \sum_{k=1}^{n_{hj}} P'_{hjk} \tag{2}$$

$$P_{1j} = 1 - \left(1 - p_{1j}\right)^{n'_1} \left(1 - p_{2j}\right)^{n'_2} \ldots \left(1 - p_{mj}\right)^{n'_m} \tag{3}$$

$$w_{jk} = \frac{1}{P'_{1j}\, P_{2jk}\, P_{3jk} \ldots} \tag{4}$$

---

*When the first-stage sampling is an unstratified sampling of schools, formulas 1 and 3 simplify to formulas 5 and 6 respectively:

$$P'_{jk} = \frac{1}{N'} \quad \text{assuming all schools have equal a priori probabilities} \tag{5}$$

$$P'_{1j} = 1 - \left(1 - \sum_{k=1}^{n_j} p'_{jk}\right)^{n'} \tag{6}$$

TABLE 1. Hypothetical joint percentage distributions for the transition from junior high school to senior high school in Community X

		% of JHS's 9th-grade enrollment												% of SHS's 10th-grade enrollment														N	% of Total
		Junior high schools												Junior high schools												Transfer in**	Total		
School* → ↓		J11	J12	J13	J14	J15	J16	J17	J18	J19	J20	J21	J22	J11	J12	J13	J14	J15	J16	J17	J18	J19	J20	J21	J22				
Senior high schools	S1	-	93	-	86	-	85	-	-	24	-	-	-	-	32	-	22	-	27	-	-	7	-	-	-	12	100	338	25
	S2	-	-	-	-	-	-	-	-	-	65	-	-	-	-	-	-	-	-	-	-	-	96	-	-	4	100	132	10
	S3	-	-	45	-	8	-	-	-	-	-	-	-	-	-	85	-	11	-	-	-	-	-	-	-	4	100	74	6
	S4	88	-	-	-	-	-	-	-	-	-	-	80	49	-	-	-	-	-	-	-	-	-	-	36	15	100	206	15
	S5	-	-	24	-	25	-	96	-	-	-	6	-	-	-	17	-	13	-	62	-	-	-	2	-	6	100	193	14
	S6	-	-	25	-	-	-	-	-	-	-	80	-	-	-	35	-	-	-	-	-	-	-	57	-	8	100	99	7
	S7	-	-	-	-	-	-	-	-	70	-	-	-	-	-	-	-	-	-	-	-	85	-	-	-	15	100	79	6
	S8	-	-	-	-	60	-	-	-	-	-	9	-	-	-	-	-	88	-	-	-	-	-	9	-	3	100	67	5
	S9	-	-	-	-	-	-	-	92	-	-	-	-	-	-	-	-	-	-	-	90	-	-	-	-	10	100	87	7
	S10	-	-	-	-	-	-	-	-	-	30	-	-	-	-	-	-	-	-	-	-	-	92	-	-	8	100	64	5
Transfer away** or drop out		12	7	6	14	7	15	4	8	6	5	5	20																
Total		100	100	100	100	100	100	100	100	100	100	100	100															1339	100
N		115	116	140	85	98	110	124	85	96	196	70	92														1327	1212***	
% of Total		9	9	11	6	7	8	10	6	7	15	5	7														100		

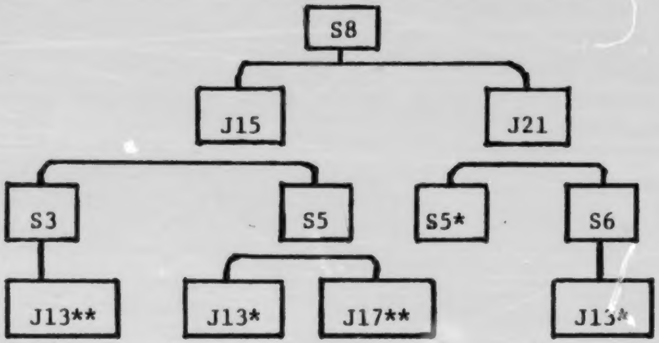
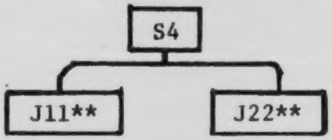
*S1, S2, ... S10 are the 10 senior high schools; J11, J12, ... J22 are the 12 junior high schools.
**Change of residence
***Total excluding transfers and dropouts

D2-10

TABLE 2. Hypothetical 1st-stage sampling to select 3 potential nucleus schools in Community X

(n´ = 3)

Pool from which nucleus schools are selected:
All senior high schools (S1, S2, ... S10)
(N´ = 10)

Order of selection	Potential nucleus school selected by random sampling	Sequential identification of schools belonging to cluster	List of schools in cluster	Cluster #
1	S8	 S8 J15 J21 S3 S5 S5* S6 J13** J13* J17** J13*	S3 J13 S5 J15 S6 J17 S8 J21	1
2	S5	***		
3	S4		S4 J11 J22	2
****				

*Same school appears to left, in same row of chart.

**Dead end; all corresponding schools have already been identified as belonging to the cluster.

***No cluster formed; selected school was rejected as nucleus school because it was in a previously selected cluster.

****No additional school selected, since it is hypothesized that it was decided to select only 3 schools from Community X as potential nucleus schools. Note that the fact that one of them (S5) did not turn out to be a nucleus school does not modify this decision.

TABLE 3. Hypothetical 2nd-stage sampling, and determination of weights for cases in the 2 selected clusters

Cluster # (j)	School (k)	Probability of selection as a nucleus school (p'_{jk})	Probability of selection (P'_{ij})	2nd-stage sampling ratio* (p_{2jk})	Weight for each student in sample (w_{jk})
1	S3	.1	.784	1/3	3.83
	S5	.1	.784	1/3	3.83
	S6	.1	.784	1/3	3.83
	S8	.1	.784	1/3	3.83
	J13	0	.784	1/3	3.83
	J15	0	.784	1/3	3.83
	J17	0	.784	1/3	3.83
	J21	0	.784	1/3	3.83
2	S4	.1	.271	1	3.69
	J11	0	.271	1	3.69
	J22	0	.271	1	3.69

---

*Assume that these values are selected as second stage sampling ratios because they are convenient to work with (i.e., selecting every 3rd student or every student) and have the desirable feature of making the weights nearly equal.

TABLE 4. Complete set of closed clusters for hypothetical data of Community X

Cluster #	Schools JHS	Schools SHS	Type of cluster*
1	J13 J15 J17 J21	S3 S5 S6 S8	D
2	J11 J22	S4	B
3	J20	S2 S10	C
4	J18	S9	A
5	J12 J14 J16 J19	S1 S7	D

---

*Definition of closed-cluster types

Type	No. of JHS's	No. of SHS's
A	1	1
B	>1	1
C	1	>1
D	>1	>1

Appendix E1

WHEN IS BIAS BETTER?*

Marion F. Shaycoft
American Institute for Research
and
University of Pittsburgh

The answer to the question, "When is bias better?" is "Not very often, but perhaps oftener than we sometimes think" -- an ambiguous answer to an ambiguous question. Let's clarify.

Suppose we are trying to estimate a population mean from a sample in which certain sizable segments of the population are represented so sparsely that to get an unbiased estimate of the population mean the comparatively few cases representing these under-represented segments have to be very heavily weighted in comparison with the other cases. Since a mean based on the under-represented segments is likely to have a larger standard error than a mean based on the better represented segments, weighting the former cases heavily could inflate the overall standard error enormously.

Giving each case in the sample equal weight, on the other hand, would generally not inflate this measure of random sampling error -- in other words the standard error of the mean -- but it would yield a biased estimate of the mean. And leaving the under-represented segments of the population out of the mean altogether, in other words giving them zero weight, would produce even more bias.

There are at least two sets of circumstances in which the available sample under-represents certain segments of the population. The first is deliberate under-representation, produced by differential sampling ratios, which it may have been necessary or desirable to use because of differential cost factors or for other very good reasons that

---

* Paper presented at the convention of the American Psychological Association, Los Angeles, September 8, 1964.

might outweigh any resulting statistical problems. The second set of circumstances is involuntary under-representation of certain segments, when one is dependent for one's data on questionnaire replies. Here the question is, "What do you do about non-respondents?" One thing you can do is to ignore them, tacitly assuming the respondents to be representative of the entire group; but this is usually an unsound assumption. Or you can take some small random sample of the non-respondents -- as large a sample as your budgetary resources will permit -- and track them to their lairs, where you will use whatever legal and ethical persuasive means are at your disposal to wring answers out of them. This is the "intensive field follow-up" method, and you can get virtually 100 percent response with it, for the cases included. But needless to say, the cost per case is high -- a very potent restriction on the proportion of non-respondents that can be included in this special-treatment group. It is certainly cheaper to eschew this method altogether, and try to get a reasonably good estimate of the non-respondents' mean by statistical manipulation of the data for respondents, considered in conjunction with whatever other data may already be available concerning the non-respondents.

One way of doing this is the method we shall call "stratification of respondents". This method requires that data on a suitable variable, which will serve as a basis for stratification, be available for both respondents and non-respondents. Let's call the variable for which data concerning the non-respondents are missing by virtue of their non-response the "dependent variable". To obtain an estimate of the non-respondents' mean on the dependent variable, the respondents are stratified on whatever stratification variable has been chosen as suitable, and a weighted average of their mean scores on the dependent variable within stratum on the stratification variable is obtained. The stratum means are weighted proportionally to the representation of the non-respondents in the various strata. (This is Method C on page 1 of the handout.) This method gives a biased estimate, but selection of a suitable stratification variable can go a long way towards reducing the bias.

The rest of this paper is concerned with comparison of the intensive follow-up method and the stratification-of-respondents method. The primary concern of the study thus is the problem of non-response to questionnaires, rather than the deliberate use of differential sampling ratios, although some of the considerations are the same in the two situations.

Two versions of the intensive field follow-up method are considered. They differ only slightly, one of them using a stratified random sample of non-respondents, and the other a simple random sample. (They are identified as Methods A and B respectively on page 1 of your handout.)

I might mention at this point that the entire theoretical development is based on the assumption of finite populations. This assumption seemed desirable since under certain circumstances the sample may be a large enough proportion of the population to invalidate the formulas that assume an infinite population.

Project TALENT is the source of the empirical data. It provides a vast body of data available about each student in it -- not just those who respond to our follow-up questionnaires after they get out of high school. Grade 12 students were used. Of the variables that were available for all cases, nine were selected for use as stratification variables and seven others for use as dependent variables. These 16 variables are shown in Table 1. The dependent variables were in effect synthetic questionnaire-item variables; in other words in the stratification-of-respondents method they were treated as if they were available only for those grade 12 students in Project TALENT who actually responded to the regular follow-up questionnaire sent out one year later. And the intensive field-follow-ups of non-respondents were also synthetic, since after we picked our random samples of non-respondents, all we had to do was assume 100 percent success in tracking them down, and then analyze the dependent variable data we already had for them. Dependent variable data based on *all* non-respondents served as a criterion against which all methods could be compared.

Of the two intensive follow-up methods, Method A and Method B, Method A was the one we tried out empirically, by actually dividing the non-respondent group into many equal-sized samples and determining the standard deviations of the sample means. This was done for samples of four different sizes ($1\frac{1}{4}\%$, $6\frac{1}{4}\%$, $12\frac{1}{2}\%$, and 25%). The Method B data were obtained primarily from theoretical formulas rather than from empirical distribution data.

A word about how Method A differs from Method B is in order at this point. In Method A, the sample is selected on the basis of terminal digits of the six-digit student testing numbers used in Project TALENT. These numbers were assigned in blocks to the schools, which were arranged by city within state, with the states arranged by geographical region. Thus selection on the basis of student number meant, in effect, that our samples were stratified by school, as well as by all the geographical and other factors that went along with it -- and even, in most cases, by home room within school since that is the way the battery was generally administered. In the case of a stratified random sample, not only is the probability of inclusion in the sample proportional to the frequency of the stratification variable category in the population, . but *actual frequencies* in these categories in the sample are proportional to the corresponding frequencies in the population. And since very wide differences have been found among the schools in regard to student performance, Method A might be expected to give somewhat more accurate estimates than Method B. I'll have more to say about that later.

As for our population, for purposes of this study we started with a group consisting of about 60,000 grade 12 students for whom the data had been put on tape and to whom follow-up questionnaires had been sent. For some complicated reasons that have nothing to do with the purposes of this study, a stratified random 20 percent of the cases were eliminated from the population, on the basis of last digit of student testing number.*

The various groups are summarized in Table 2 in terms of number of cases and mean scores. It is quite clear from this table that elimination of 20 percent of the cases has left the distributions virtually unaltered but that respondents are by no means representative of the total group, having markedly higher means than the non-respondents. Table 3 summarizes these differences.

Table 4 shows the correlations of the response-nonresponse dichotomy with the stratification variables and the dependent variables, and also the correlations between the stratification and dependent variables. The latter correlations are also shown based just on the respondent group -- which is the only way they would normally be available operationally. The selection factor among respondents is apparently not large enough to affect intercorrelations markedly. It can be seen from Table 4 that r'_{UX} (the correlation based on respondents only) is about the same as r_{UX} (the correlation based on the total group, including nonrespondents). Most of the correlations with response-nonresponse are about .25 or lower.

---

*All 62,716 cases were used, however, as a basis for the augmented stanines on the stratification variables. These stanines constituted the strata used in the stratification-of-respondents method. An augmented stanine scale is like a conventional stanine scale except that it has eleven intervals (designated 0-10) instead of nine. There is one extra interval at each end, cutting off the top and bottom 1.22 %.

The best stratification variable isn't necessarily the one that correlates highest with the dependent variable, assuming, of course, that all correlations fall short of unity. (Usually, however, the best stratification variable does correlate highest.) Nor is it necessarily the one with the highest correlation with the response-nonresponse dichotomy. The best stratification variable is one for which the partial correlation $r_{XV.U}$ is closest to 0 where U is the stratification variable, partialed out of the correlation between the dependent variable and response-nonresponse.

The partial correlation coefficients are shown in Table 4. For each dependent variable the stratification variable that appears best in the sense that it yields the partial correlation closest to zero has been checked at the right, and these pairs of variables have been recapitulated in Table 5.

The best stratification variable turns out to be the same one for girls as for boys in every case except the one for X4, Mechanical Information, for which U6, Electrical and Electronic Information, gives the best results for boys while for girls U3, Reading Comprehension, is best. This discrepancy is almost certainly due to the fact that most girls score very low on Electrical and Electronic Information.

Formula 12, on page 5 of the handout, might under certain circumstances constitute an alternative to the use of $r_{XV.U}$ as a means of selecting a stratification variable. This formula provides a way of estimating the unknown correlation between dependent variable and stratification variable based on the entire group, both respondents and nonrespondents, assuming the stratification variable is an ideal one (in the sense of producing a zero partial correlation, $r_{XV.U}$). It is also assumed, in the derivation of formula 12, that $r_{UX.V} = r'_{UX}$. In case any fragmentary data are available that lead one to suspect that the

r_{UX} estimate yielded by formula 12 is close to the actual correlation for one of the potential stratification variables and quite far from it for the others, this may be helpful in selecting the stratification variable to be used operationally, although we have no empirical evidence to this effect. As a matter of fact, while formula 12 seems theoretically sound, being based on what I regard as tenable assumptions, it turns out to be fairly insensitive, so that it wouldn't be of much practical use operationally unless the correlations of the respondent-nonrespondent dichotomy with the stratification and dependent variables were very high -- in other words, unless respondent bias were terrible. (I'm happy to say, incidentally, that when formula 12 was tried out empirically on the Project TALENT follow-up data it didn't work at all. It was of no use in helping to select stratification variables, since the respondent bias is apparently not sufficiently bad, although it does exist. According to the formula 12 results, all potential stratification variables were fine for the Project TALENT data.)

Table 6 is set up with three horizontal bands of data (one for each of the three methods) and with a two-column vertical strip for each dependent variable. In each case the numbers in the second of the two columns represent error in the estimate of the mean. The population of respondents is defined as Group A1 (see Table 2) and the population of non-respondents as Group A2. For methods A and B, in which only random error is involved (since they give unbiased estimates), the error measure shown is the standard error of the mean. For method C the measure of systematic error (bias) is shown. Because the group used is considered the population, systematic error is the only error in the means estimated by Method C. The sampling error vanishes since the estimates are based on the entire population. Thus the standard errors of the means shown for methods A and B, and measures of bias shown for Method C are in the same metric and are directly comparable. A comparison of these values for any X variable gives us an idea of which method is doing the best job.

It will be seen from Table 6 that the basis used for determining the best stratification variable (i.e. how close $r_{XV.U}$ is to 0) works astonishingly well. In the bottom section of Table 6, where the results for the stratification-of-respondents method (Method C) are shown, a number symbol (#) has been put after the measurement of bias for the theoretically best stratification variable. And in every case, no matter how scant the margin by which it is theoretically best, this variable gives the best results, in the sense of the least bias.

A direct comparison of the standard errors of the means for Methods A and B shows no marked superiority of the former over the latter. Frankly, I'm just a little puzzled as to why not. There may be some artifact creeping in because the Method B results were determined by formula instead of being based on empirical data from actual random samples. That's a rather weak hypothesis but it's the best I have available at the moment. But that is a side issue. The important problem under consideration here is not comparison of two unbiased estimates, but comparison of an unbiased estimate based on a relatively small sample with a biased estimate based on a larger sample.

In regard to this basic problem of comparing stratification of respondents with an intensive follow-up of a nonrespondent sample, the picture differs from variable to variable. For some dependent variables there is one outstandingly good stratification variable that gives results about as good as a 5 percent intensive follow-up of nonrespondents would. And for at least on X variable, whether the student plans to major in math, use of the best stratification variable, the Mathematics Information score, gives better results than even a 25% follow-up of nonrespondents. But in some cases even the best available stratification variable gives poorer results than a mere $1\frac{1}{4}$ percent sample of the non-respondents. And in hardly any instance does more than one of the stratification variables work out very well for the same dependent variable.

In fact some of the more inappropriate stratification variables are worse than none at all. They give a more biased estimate than would result from the assumption that respondents are representative of non-respondents. Furthermore it must be borne in mind that even though one of the stratification variables works very well, it may not be possible to identify that variable in the operational situation. The principle that $r_{XV.U}$ should be near zero for a good stratification variable works admirably, but in operational use it might not be possible to estimate $r_{XV.U}$ accurately since the correlation between U and X could be determined only for respondents and r_{VX} would usually be totally unknown. Thus intensive follow-up of a comparatively small sample of respondents -- 5%, perhaps -- is probably the safest procedure. The amount of sampling error that is likely to occur with any given sample size can be estimated reasonably well in advance and one is guarded against a bad choice of stratification variable.

On the opposite side of the picture, however, is the fact that familiarity with the variables one is using and shrewd judgment may help one to pick a good stratification variable. Furthermore, stratification variables better than any of those tried in the present study could almost certainly be obtained by using an optimally weighted composite of scores on several tests.

In summary, then, although there are arguments on both sides, the intensive field follow-up procedure is generally best (at least in a large scale study) when the follow-up sample is as large as 5 percent of the nonrespondents. However, where this isn't possible, reasonably good results can probably be obtained by stratification of respondents, if considerable effort is devoted to determining the best stratification variable available.

This study has been concerned with the effect of non-response on estimates of the means. Further research should probably be devoted to the effect on estimates of the standard deviation and other important statistics.

E1-10

Formulas and Tables

to accompany

When is Bias Better?

Marion F. Shaycoft
American Institute for Research
and
University of Pittsburgh

NOTATION

N = number of cases in finite population (of non-respondents)

Variables

X = variable for which the mean (for non-respondents) is to be estimated.

U = variable used as a basis for stratification of respondents. This stratification variable must be one for which scores are known for non-respondents as well as for respondents.

V = dichotomous variable representing whether the individual responds to the questionnaire.

μ = mean on Variable X for population of non-respondents

Methods of estimating μ

Method A
Intensive follow-up of comparatively small stratified random sample of non-respondents.

Method B
Intensive follow-up of comparatively small simple random sample of non-respondents.

Method C
Stratification of respondents on a suitable variable. To obtain an estimate of the non-respondents' mean score on the dependent variable, a weighted average of the respondents' mean scores within stratum is obtained, weighting proportionally to the representation of the non-respondents in the various strata.

Statistics used in Method A

m = number of samples of non-respondents.

n_i = number of cases in non-respondent sample i. $(i = 1,2,\ldots m)$

$$\sum_{i=1}^{m} n_i = N$$

$\bar{n}_i$ = mean number of cases in sample

$$\bar{n}_i = \frac{N}{m}$$

M_i = unbiased estimate of μ, based on sample i.

σ_{M_i} = root-mean-square of deviations of M_i from μ.

Thus $\sigma_{M_i}^2$ = unbiased estimate of variance error of M_i

Statistics used in Method B

n = number of cases in simple random sample of non-respondents.

σ^2 = population variance of scores on Variable X.

$\hat{\sigma}^2$ = unbiased estimate of σ^2.

M = mean of Variable X, based on simple random sample of n cases.

σ_M^2 = unbiased estimate of variance error of $M = E(M - \mu)^2$.

σ_M = corresponding standard error of M.

p = sampling ratio = $\frac{n}{N}$

$q = 1 - p$

Statistics used in Method C

$\overline{X}'$ = biased estimate of μ, computed by stratifying respondents on Variable U, and weighting stratum means (for Variable X) differentially.

n_j = number of non-respondents in stratum j on Variable U.

$$\sum n_j = N$$

n'_j = number of respondents in stratum j on Variable U.

X_{jk} = score on Variable X for respondent k in stratum j on Variable U.

$$(k = 1, 2, \ldots n'_j)$$

$\overline{X}_j$ = mean score on Variable X for respondents in stratum j on Variable U.

$\overline{X}$ = $E(\overline{X}')$ = expectation of $\overline{X}'$.

k = amount of bias (constant error).

$$k = \overline{X} - \mu$$

$\sigma_{\overline{X}'}$ = standard error of $\overline{X}'$ (= amount of random error in $\overline{X}'$)

$$\sigma_{\overline{X}'}^2 = E(\overline{X}' - \overline{X})^2$$

$\epsilon_{\overline{X}'}$ = measure of total error in $\overline{X}'$

$$\epsilon_{\overline{X}'}^2 = E(\overline{X}' - \mu)^2$$

r'_{UX} = correlation between Variables U and X, based on respondents only.

FORMULAS

Method A

$$\mu = E(M_i) = \frac{\sum_{i=1}^{m} n_i M_i}{N} \tag{1}$$

$$\sigma_{M_i}^2 = \frac{\sum_{i=1}^{m} n_i (M_i - \mu)^2}{N} \tag{2}$$

$$= \frac{\sum n_i M_i^2}{N} - \left(\frac{\sum n_i M_i}{N}\right)^2 \tag{3}$$

Method B

$$\hat{\sigma}^2 = \frac{\sum_{1}^{n} (X - \mu)^2}{n} = \frac{\sum (X - M)^2}{n - 1} \cdot \frac{N-1}{N} \tag{4}$$

$$\text{If } n = N : \quad \hat{\sigma}^2 = \frac{\sum_{1}^{N} (X - \mu)^2}{N} = \sigma^2 \tag{5}$$

$$\sigma_M^2 = \frac{N-n}{n(N-1)} \cdot \hat{\sigma}^2 = \frac{q}{p\,(N-1)} \cdot \hat{\sigma}^2 \tag{6}$$

Method C

$$E(\bar{X}') = \bar{X}' \qquad (7)$$

$$\epsilon_{\bar{x}'} = \sqrt{\sigma_{\bar{x}'}^2 + k^2} \qquad (8)$$

When $\sigma_{\bar{x}'} = 0$:

$$\epsilon_{\bar{x}'} = k \qquad (9)$$

$$\bar{X}_j = \frac{\sum_{k=1}^{n_j'} X_{jk}}{n_j'} \qquad (10)$$

$$\bar{X} = \frac{\sum n_j \bar{X}_j}{\sum n_j} = \frac{\sum n_j \bar{X}_j}{N} \qquad (11)$$

Estimated r_{UX} assuming U is an ideal stratification variable ;

$$r_{UX_{est.}} = \frac{r'_{UX}}{\sqrt{1 - r_{UV}^2 \left(1 - r'^2_{UX}\right)}} \qquad (12)$$

Assumptions underlying formula 12 :

For ~~any~~ strat. variable: $r_{UX.V} = r'_{UX}$ (13)

For an ideal strat. variable: $r_{XV.U} = 0$ (14)

Best method

The best method is defined as the one that gives the lowest value of ϵ where ϵ is defined as in formula 8. (If $\bar{X}'$ happens to be an unbiased estimate of the mean, k = 0 and $\epsilon = \sigma_M$.) (15)

Table 1. Summary of variables used

Category of variables	Source	Var. #	Var. Code	Description of variable	Kind of scale
Dependent variables	Test Scores*	X1	R-312	Math II. Introductory h. s. math.	Raw score
		X2	R-333	Math III. Advanced h. s. math.	" "
		X3	R-230	English	" "
		X4	R-112	Mechanical Information	" "
		X5	R-281	Visualization in 2 Dimensions	" "
	Question-naire**	X6	SIB-303-B	Plan to start college right after h. s.?	Percent
		X7	SIB-210-E	Plan to major in math?	"
Stratification variables	Test Scores*	U1	R-106	Math Information	Augmented stanine
		U2	R-311	Math I. Arithmetic Reasoning	" "
		U3	R-250	Reading Comprehension	" "
		U4	R-290	Abstract Reasoning	" "
		U5	R-212	Memory for Words	" "
		U6	R-111	Electricity and Electronics Info.	" "
		U7	R-270	Mechanical Reasoning	" "
		U8	R-282	Visualization in 3 Dimensions	" "
		U9	F-410	Arithmetic Computation	" "

* Tests in the Project TALENT Battery

** Options of items in the Project TALENT Student Information Blank

Table 2. inition of groups, and comparison among group means on dependent variables

Sex	Group Designation	Description*	Terminal digit of student testing#	Number of cases	Mean**** or S D.	X1	X2	X3	X4	X5	X6	X7
M	A	Group used in this study (80% of total)	1-4 6-9									
	A1	Respondents***		15143	M	13.62	5.11	85.29	13.79	14.99	57.25	3.30
	A2	Non-respondents***		8940	M	10.36	3.50	78.29	12.67	13.56	34.26	1.87
	A	Total		24083	M	12.41	4.51	82.69	13.37	14.46	48.71	2.77
	B	Group not used (20% of total)	0,5	6082								
	T	Grand total**		30165	M	12.41	4.52	82.70	13.37	14.46	48.82	2.77
		" "		"	σ	5.69	3.17	13.81	3.32	5.74	49.99	16.41
F	A	Group used in this study (80% of total)	1-4 6-9									
	A1	Respondents***		17741	M	11.05	3.45	90.50	8.60	12.35	50.10	1.71
	A2	Non-respondents***		8254	M	8.80	2.65	85.00	7.59	11.01	30.63	0.68
	A	Total		25995	M	10.34	3.20	88.75	8.28	11.92	43.92	1.38
	B	Group not used (20% of total)	0,5	6556								
	T	Grand total**		32551	M	10.34	3.20	88.73	8.28	11.92	43.89	1.35
		" "		"	σ	4.89	2.41	12.47	3.08	5.64	49.63	11.54

* All cases are Project TALENT students who were in grade 12 in the spring of 1960.

** The complete group consists of all grade 12 students on the 1963 edition of the Project TALENT master tape.

*** To the follow-up questionnaire sent in the spring of 1961.

**** Neither these means nor any of the other data in this paper are to be regarded as normative data, since each case has been given unit weight instead of being weighted to adjust for different effective sampling ratios.

Table 3. Respondent bias

Var. #	Var. code	Description of variable	Difference between means: Respondent minus total (Group A1-A) Boys	Girls	Respondent minus non-respondent (Group A1-A2) Boys	Girls
X1	R-312	Math II	1.21	.71	3.26	2.25
X2	R-333	Math III	.60	.25	1.61	.80
X3	R-230	English	2.60	1.75	7.00	5.50
X4	R-112	Mech. Info.	.42	.32	1.12	1.01
X5	R-281	Vis. 2 D.	.53	.43	1.43	1.34
X6	SIB-303-B	Plan college	8.54	6.18	22.99	19.47
X7	SIB-210-E	Plan math major	.53	.33	1.43	1.03

Table 4. Statistics relevant to selection of stratification variables.

Based on grade 12 students (Group A)

Boys: N = 24083*
Girls: N = 25995*

	BOYS						GIRLS						Best Strat.	
Dep. var.	r_{XV}^{**}	Strat. var.	r_{UV}^{**}	r_{UX}	r'_{UX}	$r_{XV.U}$	r_{XV}^{**}	Strat. var.	r_{UV}^{**}	r_{UX}	r'_{UX}	$r_{XV.U}$	Boys	Girls
X1	.277	U1	.280	.805	.815	.091	.214	U1	.216	.747	.767	.081	✓	✓
		U2	.244	.709	.704	.152		U2	.201	.631	.633	.115		
		U3	.264	.636	.627	.147		U3	.230	.591	.589	.099		
		U4	.213	.552	.541	.196		U4	.180	.495	.486	.146		
		U5	.140	.418	.400	.243		U5	.132	.414	.408	.177		
		U6	.221	.484	.469	.199		U6	.138	.361	.367	.178		
		U7	.195	.480	.469	.213		U7	.176	.451	.448	.153		
		U8	.174	.441	.435	.227		U8	.156	.409	.409	.166		
		U9	.194	.561	.548	.207		U9	.148	.510	.508	.163		
X2	.246	U1	.280	.733	.750	.062	.154	U1	.216	.616	.647	.027	✓	✓
		U2	.244	.593	.599	.130		U2	.201	.452	.473	.072		
		U3	.264	.547	.550	.126		U3	.230	.433	.449	.062		
		U4	.213	.484	.483	.167		U4	.180	.371	.382	.095		
		U5	.140	.359	.353	.212		U5	.132	.299	.306	.121		
		U6	.221	.438	.432	.170		U6	.138	.315	.328	.118		
		U7	.195	.414	.409	.185		U7	.176	.361	.375	.099		
		U8	.174	.392	.392	.196		U8	.156	.329	.343	.110		
		U9	.194	.449	.447	.181		U9	.148	.336	.352	.112		
X3	.243	U1	.280	.579	.585	.103	.205	U1	.216	.482	.493	.118		
		U2	.244	.632	.611	.118		U2	.201	.604	.602	.107		
		U3	.264	.637	.634	.101		U3	.230	.611	.607	.084	✓	✓
		U4	.213	.473	.460	.165		U4	.180	.472	.455	.138		
		U5	.140	.463	.442	.203		U5	.132	.499	.490	.162		
		U6	.221	.371	.350	.178		U6	.138	.265	.261	.176		
		U7	.195	.404	.380	.183		U7	.176	.391	.378	.150		
		U8	.174	.338	.320	.199		U8	.156	.339	.325	.164		
		U9	.194	.506	.480	.171		U9	.148	.497	.483	.153		
X4	.163	U1	.280	.391	.364	.061	.153	U1	.216	.378	.369	.079		
		U2	.244	.416	.383	.070		U2	.201	.402	.394	.080		
		U3	.264	.419	.386	.060		U3	.230	.429	.414	.062		✓
		U4	.213	.325	.294	.101		U4	.180	.314	.299	.103		
		U5	.140	.193	.163	.140		U5	.132	.236	.226	.127		
		U6	.221	.616	.599	.035		U6	.138	.438	.441	.104	✓	
		U7	.195	.535	.524	.071		U7	.176	.440	.434	.085		
		U8	.174	.399	.385	.104		U8	.156	.365	.365	.104		
		U9	.194	.198	.157	.130		U9	.148	.243	.225	.122		
X5	.121	U1	.280	.310	.296	.037	.110	U1	.216	.277	.270	.053		
		U2	.244	.320	.304	.047		U2	.201	.318	.303	.050		
		U3	.264	.322	.290	.039		U3	.230	.327	.302	.038		
		U4	.213	.400	.374	.040		U4	.180	.398	.384	.043		
		U5	.140	.185	.171	.098		U5	.132	.215	.202	.084		
		U6	.221	.293	.280	.060		U6	.138	.184	.179	.087		
		U7	.195	.489	.474	.030		U7	.176	.450	.443	.035	✓	✓
		U8	.174	.483	.471	.043		U8	.156	.464	.456	.043		
		U9	.194	.234	.219	.079		U9	.148	.248	.235	.077		
X6	.221	U1	.280	.482	.479	.102	.183	U1	.216	.452	.455	.098	✓	✓
		U2	.244	.355	.350	.148		U2	.201	.290	.292	.133		
		U3	.264	.411	.406	.128		U3	.230	.364	.361	.110		
		U4	.213	.295	.281	.169		U4	.180	.260	.254	.143		
		U5	.140	.258	.255	.193		U5	.132	.237	.236	.158		
		U6	.221	.233	.223	.179		U6	.138	.189	.186	.161		
		U7	.195	.212	.198	.187		U7	.176	.230	.232	.149		
		U8	.174	.185	.170	.195		U8	.156	.182	.182	.159		
		U9	.194	.328	.320	.170		U9	.148	.219	.213	.156		
X7	.039	U1	.280	.128	.134	.003	.037	U1	.216	.180	.193	-.002	✓	✓
		U2	.244	.096	.104	.016		U2	.201	.116	.126	.014		
		U3	.264	.048	.045	.027		U3	.230	.084	.088	.018		
		U4	.213	.075	.075	.024		U4	.180	.100	.108	.019		
		U5	.140	.058	.058	.031		U5	.132	.055	.060	.030		
		U6	.221	.020	.012	.035		U6	.138	.076	.082	.027		
		U7	.195	.041	.040	.031		U7	.176	.094	.104	.021		
		U8	.174	.055	.051	.030		U8	.156	.094	.103	.023		
		U9	.194	.104	.109	.019		U9	.148	.115	.131	.020		

---

* r'_{UX} is based on 15143 male respondents or 17741 female respondents.

** r_{XV} and r_{UV} are point biserial correlations.

Table 5. Best* stratification variable corresponding to each dependent variable.

Dependent variable			Best strat. variable for boys			Best strat. variable for girls		
X1	R-312	Math II	U1	R-106	Math Info.		same	
X2	R-333	Math III	U1	R-106	Math Info.		same	
X3	R-230	English	U3	R-250	Rdg. Comp.		same	
X4	R-112	Mech. Info.	U6	R-111	Elec. Info.	U3	R-250	Rdg. Comp.
X5	R-281	Vis. in 2 Dim.	U7	R-270	Mech. Reas.		same	
X6	SIB-303-B	Plan college	U1	R-106	Math. Info.		same	
X7	SIB-210-E	Plan math major	U1	R-106	Math. Info.		same	

* By the "best" stratification variable is meant the one, of the nine for which data are available, that yields the $r_{XV.U}$ closest to 0.

Table 6a. Comparison of errors of estimating means in various ways for non-respondents

Grade 12 boys

Procedure					X1 (R-312) Math II		X2 (R-333) Math III		X3 (R-230) English		X4 (R-112) Mech. Inf.		X5 (R-281) Vis. 2 Dim.		X6 (SIB-303-B) Plan college		X7 (SIB-210-E) Plan math major	
A. Intensive follow-up of strat. random sample.	% of non-respondents per sample	$\bar{n}_1$	No. of samples	No. of component 1.25% samples per sample	Range of means (M_1's)	σ_{M_1}	Range of means (M_1's)	σ_{M_1}	Range of means (M_1's)	σ_{M_1}	Range of means (M_1's)	σ_{M_1}	Range of means (M_1's)	σ_{M_1}	Range of %'s	$\sigma_{\%_1}$	Range of %'s	$\sigma_{\%_1}$
(Data based on 8940 male non-respondents.)	1.25	112	80	1	9.2-11.4	.53	3.0-4.2	.26	74.5-81.5	1.48	12.1-13.2	.26	12.1-14.4	.54	23.5-45.6	4.43	0.0-5.3	1.22
	6.25	559	16	5	9.9-10.8	.24	3.3-3.7	.10	77.3-79.3	.50	12.4-12.9	.13	13.2-14.0	.25	31.2-38.8	1.95	1.0-2.6	.45
	12.50	1118	8	10	10.0-10.7	.20	3.4-3.6	.06	77.9-78.7	.27	12.5-12.8	.10	13.3-13.9	.23	31.9-38.6	1.81	1.6-2.3	.27
	25.00	2235	4	20	10.2-10.5	.11	3.5-3.6	.03	78.0-78.5	.20	12.6-12.7	.04	13.5-13.6	.02	32.3-36.3	1.32	1.6-2.0	.17
———	100.00*	8940	1	80	μ =10.36	0	μ =3.50	0	μ =78.29	0	μ =12.67	0	μ =13.56	0	μ =34.26	0	μ =1.87	0
B. Intensive follow-up of simple random sample.	% of non-respondents in sample	n			σ	σ_M	σ	σ_M	σ	σ_M	σ	σ_M		σ_M	σ	σ_M	σ	σ_M
(Standard errors of means are computed from formula, assuming a finite population of 8940 non-respondents.)	1.00	89			Same as for 100% sample	.53	Same as for 100% sample	.28	Same as for 100% sample	1.55	Same as for 100% sample	.36	Same as for 100% sample	.63	Same as for 100% sample	4.99	Same as for 100% sample	1.43
	1.25	112				.48		.25		1.38		.32		.56		4.46		1.27
	2.50	224				.34		.17		.97		.23		.39		3.13		.90
	5.00	447				.23		.12		.68		.16		.28		2.19		.62
	6.25	559				.21		.11		.60		.14		.24		1.94		.56
	10.00	894				.16		.08		.47		.11		.19		1.51		.43
	12.50	1118				.14		.07		.41		.10		.17		1.33		.38
	20.00	1788				.11		.06		.31		.07		.13		1.00		.29
	25.00	2235				.09		.05		.27		.06		.11		.87		.25
	50.00	4470				.05		.03		.16		.04		.06		.50		.14
	100.00*	8940			5.08	0	2.62	0	14.71	0	3.44	0	5.98	0	47.46	0	13.55	0
C. Stratification	Stratification variable				$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)	$\bar{X}$	Bias** ($\bar{X}-\mu$)
(Data based on 15143 male respondents weighted to yield, est. of means for 8940 male non-respondents.)	U1 R-106 Math Inf.				10.80	.44#	3.69	.19#	80.78	2.49	13.08	.41	13.99	.43	42.40	8.14#	2.03	.16#
	U2 R-311 Ar. Reas.				11.52	1.16	4.11	.61	81.09	2.80	13.11	.44	14.08	.52	48.08	13.82	2.45	.58
	U3 R-250 Rd. Comp.				11.62	1.26	4.15	.65	80.54	2.25#	13.04	.37	14.02	.46	45.66	11.40	2.88	1.01
	U4 R-290 Abst. Reas.				12.20	1.84	4.41	.91	82.46	4.17	13.32	.65	14.00	.44	50.60	16.34	2.78	.91
	U5 R-212 Mem. Wds.				12.94	2.58	4.77	1.27	83.49	5.20	13.63	.96	14.71	1.15	53.61	19.35	3.00	1.13
	U6 R-111 Elec. Inf.				12.39	2.03	4.47	.97	83.19	4.90	12.88	.21#	14.26	.70	52.13	17.87	3.18	1.31
	U7 R-270 Mech. Reas.				12.51	2.15	4.56	1.06	83.26	4.97	13.07	.40	13.87	.31#	53.16	18.90	3.03	1.16
	U8 R-282 Vis. 3 D				12.74	2.38	4.66	1.16	83.79	5.50	13.34	.67	14.01	.45	54.24	19.98	3.01	1.14
	U9 F-410 Ar. Comp.				12.34	1.98	4.52	1.02	82.68	4.39	13.56	.89	14.50	.94	50.77	16.51	2.59	.72

* Population parameter data are shown in this row.

** In this special case, the measure of total error (∈) and bias are identical, since the population of non-respondents is defined to coincide with the total available sample of non-respondents.

# Best stratification variable, according to Table 5.

Table 6b. Comparison of errors of estimating means in various ways for non-respondents

Grade 12 girls

Procedure					X1 (R-312) Math II		X2 (R-333) Math III		X3 (R-230) English		X4 (R-112) Mech. Inf.		X5 (R-281) Vis. 2 Dim.		X6 (SIB-303-B) Plan college		X7 (SIB-210-E) Plan math major	
A. Intensive follow-up of strat. random sample. (Data based on 8254 female non-respondents.)	% of non-respondents per sample	$\bar{n}_i$	No. of samples	No. of component 1.25% samples per sample	Range of means (M_i's)	σ_{M_i}	Range of means (M_i's)	σ_{M_i}	Range of means (M_i's)	σ_{M_i}	Range of means (M_i's)	σ_{M_i}	Range of means (M_i's)	σ_{M_i}	Range of $\not{p}$'s	$\sigma_{\not{p}_i}$	Range of $\not{p}$'s	$\sigma_{\not{p}_i}$
	1.25	103	80	1	7.9-9.6	.39	2.2-3.0	.18	81.1-87.5	1.37	7.1-8.5	.27	9.7-12.7	.58	17.0-39.0	4.50	0.0-2.9	.78
	6.25	516	16	5	8.5-9.1	.14	2.5-2.8	.08	84.1-86.1	.56	7.4-7.7	.07	10.7-11.7	.23	26.2-34.3	2.05	0.2-1.1	.23
	12.50	1032	8	10	8.6-8.9	.10	2.5-2.7	.07	84.4-85.6	.40	7.5-7.7	.05	10.7-11.4	.18	28.4-31.6	1.06	0.4-0.9	.18
	25.00	2064	4	20	8.8-8.8	.02	2.6-2.7	.04	84.8-85.2	.12	7.5-7.6	.03	10.8-11.2	.15	30.4-30.9	.18	0.5-0.8	.09
	100.00*	8254	1	80	μ = 8.80	0	μ = 2.65	0	μ = 85.00	0	μ = 7.59	0	μ = 11.01	0	μ = 30.63	0	μ = 0.68	0
B. Intensive follow-up of simple random sample. (Standard errors of means are computed from formula, assuming a finite population of 8254 non-respondents.)	% of non-respondents in sample	n			σ	σ_M	σ	σ_M	σ	σ_M	σ	σ_M	σ	σ_M	σ	σ_M	σ	σ_M
	1.00	83			Same as for 100% sample	.45	Same as for 100% sample	.22	Same as for 100% sample	1.47	Same as for 100% sample	.32	Same as for 100% sample	.62	Same as for 100% sample	5.04	Same as for 100% sample	.90
	1.25	103				.41		.20		1.32		.29		.56		4.51		.80
	2.50	206				.29		.14		.93		.20		.39		3.17		.57
	5.00	413				.20		.10		.65		.14		.27		2.21		.39
	6.25	516				.18		.09		.57		.13		.24		1.97		.35
	10.00	825				.14		.07		.44		.10		.19		1.52		.27
	12.50	1032				.12		.06		.39		.09		.17		1.34		.24
	20.00	1651				.09		.04		.30		.07		.13		1.01		.18
	25.00	2064				.08		.04		.26		.06		.11		.88		.16
	50.00	4127				.05		.02		.15		.03		.06		.51		.09
	100.00*	8254			4.15	0	2.00	0	13.46	0	2.96	0	5.68	0	46.10	0	8.22	0
C. Stratification (Data based on 17741 female respondents, weighted to yield est. of means for 8254 female non respondents.)	Stratification variable				$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)	$\bar{X}$	Bias** ($\bar{X} - \mu$)
	U1 R-106 Math. Inf.				9.22	.42#	2.71	.06#	87.72	2.72	8.07	.48	11.63	.62	38.81	8.18#	0.72	.04#
	U2 R-311 Ar. Reas.				9.68	.88	2.95	.30	87.46	2.46	8.06	.47	11.60	.59	43.66	13.03	1.06	.38
	U3 R-250 Rd. Comp.				9.60	.80	2.92	.27	86.72	1.72#	7.95	.36#	11.45	.44	40.99	10.36	1.22	.54
	U4 R-290 Abst. Reas.				10.09	1.29	3.09	.44	88.23	3.23	8.22	.63	11.47	.46	44.89	14.26	1.23	.55
	U5 R-212 Mem. Wds.				10.46	1.66	3.23	.58	88.78	3.78	8.39	.80	12.02	1.01	46.73	16.10	1.49	.81
	U6 R-111 Elec. Inf.				10.48	1.68	3.20	.55	89.58	4.58	8.20	.61	12.05	1.04	47.26	16.63	1.38	.70
	U7 R-270 Mech. Reas.				10.21	1.41	3.11	.46	88.80	3.80	8.10	.51	11.40	.39#	45.77	15.14	1.27	.59
	U8 R-282 Vis. 3 D				10.39	1.59	3.18	.53	89.21	4.21	8.23	.64	11.47	.46	47.17	16.54	1.33	.65
	U9 F-410 Ar. Comp.				10.24	1.44	3.17	.52	88.64	3.64	8.37	.78	11.93	.92	46.70	16.07	1.22	.54

* Population parameter data are shown in this row.

** In this special case, the measure of total error (ϵ) and bias are identical, since the population of non-respondents is defined to coincide with the total available sample of non-respondents.

# Best stratification variable, according to Table 5.

Appendix E2

A Stratified Weights Procedure to Compensate for Nonrespondents

Marion F. Shaycoft

A. Introduction

In this appendix a description is presented of the stratified weighting scheme that is being used to correct for nonrespondent bias, along with data indicating its effectiveness. Although there are numerous alternative schemes, this scheme was developed by a careful logical and statistical rationale, and although it is algorithmically quite complex, its steps are all necessary and justified.

---

* This appendix was originally Appendix A in: McLaughlin, D. H., Fulscher, G. V., & Yen, Wendy M. Project TALENT's Special Sample: Is is Necessary? Palo Alto, Calif.: American Institutes for Research, May 1974.

B. Description of the Method

1. Overview

The general plan was to stratify the cases from whom responses were obtained, whether by mail or by interview, on appropriate variables from the original (1960) data collection, and then to apply a corrective weight to these cases, to make them represent not only themselves and the persons they already represented but also any nonrespondents to the special follow-up who happened to fall in the same stratum or cell (defined by the stratification variables) and the persons those nonrespondents already represented.

As a very simple example of the procedure, let us assume we have selected two stratification variables, U and V, and that one of them, Variable U, is a dichotomy with categories designated "No" and "Yes," while the other, Variable V, has three categories, designated "High," "Medium," and "Low" respectively. Table A-1 presents some hypothetical data, and shows how ratio R_j, a ratio

Table A-1. Hypothetical data illustrating the computation of corrective ratio R_j*.

Variable V	Variable U						Total	
	No			Yes			T	M
	T_j	M_j	R_j	T_j	M_j	R_j		
High	8,000	4,000	2.000	12,000	7,500	1.600	20,000	11,500
Medium	24,000	9,000	2.667	21,000	10,000	2.100	45,000	19,000
Low	16,000	1,000	16.000	9,000	2,000	4.500	25,000	3,000
Total	48,000	14,000		42,000	19,500		90,000	33,500

*Ratio R_j is the multiplier applied to the Weight A value of each respondent to yield his Weight C value.

which applies to the jth cell and is determined separately for each of the six cells, is computed. Ratio R_j is multiplied by Weight A (the follow-up weight that assumes that data are available for all special sample cases and therefore does not correct for nonrespondents to the special follow-up) for each respondent to determine Weight C, the stratified weight to correct for otherwise unrepresented nonrespondents. Notation and formulas are shown in the next section of this appendix.

2. Notation and formulas

The notation used in Table A-1 and throughout this appendix is as follows:

W_0 = original weight, for 1960 data

W_A = follow-up weight A

W_B = follow-up weight B

W_C = follow-up weight C

r = special sample selection ratio. For instance if every 24th nonrespondent to the mailed questionnaire is selected for inclusion in the special sample, $r = 1/24$.

k_A = follow-up ratio for computing Weight A

- $k_A = 1$ for respondents to mailed questionnaire
- $k_A = \frac{1}{r}$ for members of special sample (whether respondents or not)
- $k_A = 0$ for all other nonrespondents to mailed questionnaire, i.e. for all nonrespondents who are not selected for inclusion in the special sample.

k_C = follow-up ratio for computing Weights B and C

- $k_C = 1$ for respondents to mailed questionnaire or special follow-up
- $k_C = 0$ for cases for whom no follow-up data are available.

$T_j = \Sigma W_A$ for all cases in stratification cell j for whom $k_A > 0$

$M_j = \Sigma W_A$ for all cases in cell j for whom $k_C = 1$ (i.e. all respondents)

$R_j = T_j / M_j$

The formulas for follow-up weights are:

$W_A = k_A W_0$

$W_C = k_C R_j W_A$

W_B = the value W_C would have if there were no stratification variables (i.e. if everyone were treated as falling in one large stratum or stratification cell).

3. Selection of the stratification variables

The possibility of selecting a new set of stratification variables and developing a new set of weights for each separate use (i.e., for use in obtaining statistics on each follow-up variable) was rejected not only as being clearly out of the question on practical grounds but also as being undesirable for other reasons as well. Therefore an "all-purpose" set of stratification variables was needed--one that would be useful in obtaining data on as many of the follow-up variables as possible.

Some readers may think the best way to stratify is to have just one or two or at most three stratification variables, and to dichotomize on each or possibly trichotomize, so as to wind up with perhaps eight or nine cells. This has the advantage of comparative simplicity but to infer missing follow-up data on the basis of what the nonrespondents were like when they were in high school, we felt that inferences based on a lot of detailed data (numerous variables and finer breakdowns) would tend to produce better inferences than would cruder classifications. However convenient it might be, certainly no single variable in (or out of) the TALENT battery could conceivably be relevant to the estimation or prediction of all the criterion variables that could possibly be of concern in TALENT research. And that generalization includes within its scope not only the individual variables in the TALENT battery but also any composites obtainable by combining individual variables in accordance with a formula. No factor analysis, component analysis, discriminant function analysis, canonical correlation, or any other form of multivariate analysis would produce a magic formula that would yield a single all-purpose variable. And if in a search for the single stratification variable we were to forsake multivariate analysis and put our trust in subjective judgment, that too would fail to reveal any single all-purpose variable relevant to the providing of answers to every question answerable through TALENT data. Since no single stratification variable would serve all purposes we needed more than one stratification variable. And for similar reasons we needed more than just two or three. As a matter of fact we ended up selecting ten--three of which were natural dichotomies and seven of which were continuous variables. In order to help insure that our set of stratification variables would incorporate as wide a range of the diverse elements represented in the TALENT battery as possible, with emphasis on those that had proved to have predictive utility in the past, 28 variables from the 1960 battery were selected (subjectively), representing those elements of the battery that had had considerable predictive utility over a wide range in previous research with the TALENT data. A complete principal components analysis was carried out on these 28 variables, and on the basis of an inspection of the results nine stratification variables were selected. These were chosen in such a way that each of the first several principal components was represented by at least one variable having a very high loading on it (preferably the variable with the highest loading if that variable hadn't already been selected to represent a previous principal component). Principal components were used in this way in order to get some representation, among the stratification variables, of as much as possible of the major components of battery variance.

The nine stratification variables selected on the basis of the principal components analysis were supplemented by a tenth variable--the only follow-up variable for which we had data on everybody--

respondent and nonrespondent alike. That variable was the dichotomy created by the question "Did the person respond to the mailed questionnaire?"

The ten stratification variables are shown in Table A-2.

4. Establishing strata on each stratification variable (and "cells" at their intersections)

As has already been indicated we rejected the possibility of splitting each continuous variable into broad intervals--for instance dichotomizing or dividing into quarters. Thus, instead of just arbitrarily dichotomizing on the seven continuous variables, we used a stanine scale for each--or more precisely an "augmented stanine" scale. Augmented stanines, like regular stanines, are normalized standard scores with a mean of 5 and a standard deviation of 2; but augmented stanines divide the scale into 11 categories instead of nine, with one extra interval (designated "10") at the top and one (designated "0") at the bottom. This may seem to some readers to have provided an excessively fine breakdown since considering all ten stratification variables together (three dichotomies and seven augmented stanine scales) defines a total of 155,897,368 cells.

However it was obvious that most of these cells would be unoccupied, not just because even the very large TALENT sample wasn't large enough to provide occupants for every one of the more than 150,000,000 cells, but also and more importantly because many of them defined extremely improbable combinations of characteristics. The void cells are not a cause for concern. There is no more reason to consider that their existence interferes with the normal course of research than it would be for a researcher using IQ's to be upset because all cells representing IQ's above 180 (going all the way up to infinity!) turned out to be void.

But even though the void cells are quite innocuous in their effect on the statistics to be derived, some readers may still wonder why so fine a breakdown--a breakdown into augmented stanines--was desirable, and whether dichotomizing wouldn't have been just as satisfactory and a lot simpler. Let's dispose of the simplicity

Table A-2. Stratification variables

Code # for variable	Kind of categorization		No. of strata	Stratification variable**
-----	Dichotomy		2	Sex
D-802	"		2	High school curriculum code (academic vs. other)
R-250	Augmented stanine*		11	Reading Comprehension
R-230	"	"	11	English Total
C´004	"	"	11	Math Composite
C´005	"	"	11	Technical Info. Composite
R-270	"	"	11	Mechanical Reasoning
R-282	"	"	11	Visualization in 3 Dimensions
P*801	"	"	11	Socioeconomic index
Q3-01	Dichotomy		2	Respond to mailed questionnaire? (1 = Yes, i.e., REGULAR; 0 = No, i.e., SPECIAL)

---

*The augmented stanines used for this purpose were based on the total group for the grade (sexes combined). Augmented stanines are the same as regular stanines, except that instead of dividing the scores into nine categories (1-9), augmented stanines divide them into eleven categories, with one extra interval (designated "10") at the top and one (designated "0") at the bottom.

**The ten stratification variables are listed in order of precedence (principal variables first).

issue first. The answer to that question is that a computer doesn't really care very much. It can just as well assign cases to one of 11 groups as to one of two groups. That leaves us with the question of whether augmented stanines are *better* in some way--more useful--than a mere dichotomy, or thirds, or quarters. The answer is that although there are some activities that almost anyone can engage in with some degree of success, there are some very important activities for which the higher the level of ability the better the prognosis and for which merely having better-than-average ability is utterly inadequate. There are some jobs, for instance, in which the demands are so severe that the person who is slightly above the population mean in terms of the required abilities can't do the job at all; the person at the 90th percentile, which gives him a stanine score of 8 (the second highest possible score on a scale that runs from 1 to 9), can possibly do the work called for, but only in a hack manner, and only at a marginal level of satisfactoriness. For such jobs it may take someone who is two or two-and-a-half standard deviations above the mean in relevant skills or abilities. Stanine 9 is about two standard deviations above the mean. If we use an "augmented stanine" scale--a scale with 11 points (running from 0 to 10) instead of the conventional nine (running from 1 to 9), augmented stanine 10 is obtained by slightly more than one percent of the population. The athletes who compete in the Olympics, even the ones who don't win any gold medals, would probably be at the 10-level on augmented stanines for skill in their respective fields. Nobody gets into the Olympics merely because he is in the top quarter of his age group.

5. Representation of cases that don't match perfectly

The entire strategy of developing corrective weights by a stratification procedure requires that every cell that contains any cases at all that lack follow-up data must contain at least one case having such data. Because of the very fine breakdown with which we started it was inevitable that this requirement would not always be met on the augmented stanine breakdown. Therefore when necessary, augmented stanines were replaced by cruder breakdowns to get individual cases matched. This adjustment was introduced for one variable at a time.

The order in which stratification variable breakdowns were thus modified was in judged order of importance of the variable (starting with the least important). In case the cruder breakdown resulted in more than one matching cell for a particular case lacking follow-up data, the closest one, on the basis of the initial fine breakdown, was the one selected*. (Mahalanobis d^2 was involved in this choice.) The procedure insured that each case with missing data was automatically represented by the case or cases most like him in terms of the initial breakdown (augmented stanines) on the stratification variables—even though condensation into cruder breakdowns might have occurred on one or more variables.

C. Empirical evidence on the effectiveness of the weights

One way of determining the effectiveness of a set of follow-up weights is to determine what values it gives for follow-up cases on a variable for which one knows what the results should turn out to be if there were no missing data. This can be accomplished very easily by using a 1960 variable for which one has data even for the nonrespondents, and seeing how well results on that variable could be estimated for the total population, estimating solely on the basis of data for the questionnaire respondents. This is exactly what we have done, for a dozen 1960 variables. The results are presented in Table A-3. In studying this table it should be borne in mind that the means obtained using the three different weighting systems have the following significance in this situation.

Weight A means: These correspond to normative data. They are based on all mailed-questionnaire respondents and on everyone in the special sample, whether they responded or not. (In brief, they are based on everybody for whom the value $k_A > 0$.) Thus everyone in the TALENT sample is either included or represented.

---

*All of this involved an extremely complex algorithm, but an extremely complex computer program handles it all automatically. The program was written by Gary V. Fulscher, to whom the writer is grateful.

Table A-3. Comparison of Grade 12 means obtained by different weighting procedures, for selected variables

Category of variables	Variable #	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
			MALES					FEMALES						
			Weighted means*			Differences		Weighted Means*			Differences		Wtd. S.D.'s* (Wt. A)	
			Wt. A	Wt. B	Wt. C	B-A	C-A	Wt. A	Wt. B	Wt. C	B-A	C-A	M	F
		Test scores												
I. Stratification variables	R-230	English Total	82.08	82.77	82.18	.69	.10	88.75	89.64	88.96	.89	.21	14.64	12.56
	R-250	Reading Comp.	32.80	33.28	32.81	.48	.01	33.65	34.16	33.60	.51	-.05	10.46	10.15
II. Variables that are components of strat. variables	R-311**	Arith. Reas.	9.68	9.86	9.74	.18	.06	8.88	9.07	8.88	.19	.00	3.65	3.61
	R-106**	Math Info.	11.33	11.58	11.35	.25	.02	8.12	8.39	8.10	.27	-.02	6.15	5.25
	R-112***	Mech. Info.	13.26	13.56	13.40	.30	.14	8.41	8.64	8.52	.23	.11	3.45	3.12
III. Variables that are operationally independent of strat. variables	R-290	Abst. Reas.	9.54	9.68	9.59	.14	.05	9.28	9.47	9.33	.19	.05	3.03	3.04
	R-114	Home Ec. Info.	8.99	9.12	9.03	.13	.04	13.47	13.64	13.52	.17	.05	2.94	3.29
	R-115	Sports Info.	8.78	8.98	8.87	.20	.09	6.06	6.20	6.09	.14	.03	2.87	2.52
		Interest scores												
	P-701	Phys Sci, Engin, Math	21.54	21.66	21.53	.12	-.01	12.44	12.54	12.40	.10	-.04	8.60	7.88
	P-714	Mech.-tech.	21.14	21.22	21.17	.08	.03	8.95	9.06	9.06	.11	.11	8.17	6.31
	P-717	Labor	12.76	12.60	12.79	-.16	.03	7.83	7.94	7.94	.11	.11	7.12	5.78
		Other												
	P*801	Socioeconomic index	98.78	99.38	99.00	.60	.22	98.78	99.22	98.82	.44	.04	10.09	9.62
Approx. N (unweighted)				12,000					13,000					

---

*Using 11-yr. follow-up weights

** Component of C´004 (Math composite)

*** Component of C´005 (Technical composite)

Weight B means: These correspond to means based only on respondents, with no adjustment of weights via stratification to correct for nonrespondents to the special follow-up.

Weight C means: These are the means based on the stratified weights.

Inspection of Table A-3, particularly columns 4, 5, 9, and 10, shows that Weight C clearly gives better results (i.e. closer to the Weight A results) than does Weight B. In the 24 comparisons between the deviation of the Weight C mean from the norm estimate (Weight A mean) and the deviation of the Weight B mean from the Weight A mean, the C deviation is smaller than the B deviation 22 times and they are tied twice. B never turns out superior to A.

On the basis of these results it appears that the Weight C's are doing an effective job of correcting for missing questionnaires.

Supplementary evidence of a somewhat more theoretical nature is presented in Table A-4. To the extent that there is within-cells homogeneity with respect to a particular criterion variable the stratification procedure is effective in producing better statistics for that variable. One measure of within-cells homogeneity is the correlation ratio, η (really a multiple correlation ratio under the present circumstances since we have ten stratification variables). However to the extent that categories (cells or strata) have very few cases in them, η is spuriously high. The epsilon coefficient ε provides a correction for this. Values of both η and ε are shown in Table A-4. The epsilon values, being corrected for fineness of grouping (i.e. small cells) are population estimates. Epsilon values corrected for attenuation, which clearly would be very close to 1, would be even more important as prima facie evidence of the effectiveness of the weighting procedure. The relevant formulas are shown in the footnotes on Table A-4. Tables A-3 and A-4 provide two very different and entirely independent kinds of evidence on this point; considered together they should convince the most skeptical skeptic of the effectivenesss of the Weight C's.

TABLE A-4. Correlation ratios (and population estimates of correlation ratios) between stratification cells (independent variable) and selected dependent variables

Based on 12th-grade 11-year follow-up respondents with data on the variable (sexes combined)

Source of variable	Variable #	Variable	N (no. of cases)	n (no. of cells)	η (correlation ratio)*	ε (Pop. est. of η)*
1960		Test scores				
data	R-172	Vocab. Total	23959	22052	.988	.837
	R-192	Info. Part II Total	24177	22252	.988	.837
	R-260	Creativity	24573	22673	.980	.698
	R-290	Abstract Reasoning	24558	22658	.979	.680
		Inventory scales				
	R-610	Mature Personality	24124	22204	.978	.673
	P-701	Interest: Phys Sci, Engin., Math	24126	22206	.967	.429
	P-715	Interest: Skilled trades	24123	22204	.964	.334
11-yr	Q-3-ed-A	Amt. of education	21323	20286	.989	.742
follow-		LRG Career Group			**	
up	LRG-1	Engin., Phys Sci, Math, Arch.	25030	22952	.980	.723
	LRG-2	Medical & Biological Sciences	"	"	.970	.537
	LRG-3	Business Administration	"	"	.967	.467
	LRG-4	Gen. Teaching & Soc. Services	"	"	.969	.515
	LRG-5	Humanities, Law, Soc. & Behav. Sci	"	"	.974	.618
	LRG-8	Business, Sales	"	"	.963	.354
	LRG-9	Mechanics, Industrial Trades	"	"	.965	.414
	LRG-10	Construction Trades	"	"	.969	.515

---

*The formulas are:

$$\eta_{xu} = r_{x\bar{x}_j} = \frac{s_{\bar{x}_j}}{s_x}$$

s_x = the standard deviation on variable X for the total sample

$\bar{X}_j$ = the mean on variable X within cell j. ($\bar{X}_j$ is the dependent variable.)

$s_{\bar{x}_j}$ = the sample standard deviation of $\bar{X}_j$

U is the basis of stratification (i.e., set of stratification variables). (U is the independent variable.)

$$\varepsilon_{xu} = \sqrt{1 - \frac{N-1}{N-n}\left(1 - \eta^2_{xu}\right)}$$

where N = no. of cases
n = no. of cells

$$r_{\bar{x}\bar{x}} = \frac{r_{xx} - \varepsilon^2}{1 - \varepsilon^2}$$

$$\varepsilon_{\infty} = \frac{\sqrt{\varepsilon^2(1-\varepsilon^2)}}{\sqrt{r_{xx}(r_{xx}-\varepsilon^2)}}$$

**The values below the double asterisk in the η column are point biserial correlation ratios.

***Split-half reliability corrected by Angoff formula 16 and adjusted for range.

And as has been demonstrated in the main part of the report the Weight C procedure outlined here would be usable and useful even if the special sample stage were skipped altogether. This results from the fact that Weight C is a procedure for providing indirect representation for those TALENT sample members who are neither respondents nor directly represented by respondents (whether "regular" or "special"). Thus the procedure is available even if all the nonrespondents to the regular questionnaire retain their nonrespondent status to the end as a result of the nonexistence of a special sample phase.

As for this potential (and in the case of the present report, actual) use of Weight C to compensate for all nonrespondents to the mailed questionnaire, the statement above is merely intended to imply that if the "special-sample" phase is omitted, it is better to try to compensate at least partially by using Weight C than to make no effort to weight in such a way as to compensate for the nonrespondents; it is not intended to imply here either that it is desirable or that it is undesirable to omit the special sample phase in future follow-ups. The reader is referred to the main part of the report for a discussion of that issue and the presentation of recommendations.

MINIMIZING TOTAL ERROR: A TRADE-OFF PROBLEM*

Marion F. Shaycoft, American Institutes for Research
Lauress L. Wise, American Institutes for Research
Donald H. McLaughlin, American Institutes for Research

A. THE PROBLEM

1. Statement of the problem in general terms

An unbiased estimate of a mean is not necessarily preferable to a biased estimate. Sometimes a biased estimate, even though by definition it involves a systematic error, may nevertheless involve a much smaller total error than the unbiased estimate would. In general terms, the problem with which this paper is concerned is how to cope with the dilemma posed by having to choose between the disadvantages of a sizable systematic error and the disadvantages of a sizable random error.

2. A hypothetical (and extreme) example

Let's consider a very extreme (and improbable) example. Suppose someone wants to estimate the average height of American 25-year-olds. Now suppose that to do this he determines the heights of a 10,000-case random sample of 25-year-old men, and a two-case (!) sample of 25-year-old women. He is very careful to select these two women randomly. What is the best way for him to combine his data to get the best estimate of the overall average?

There are innumerable ways our investigator could get a biased average but there is essentially only one way to get an unbiased average that utilizes all available data. The unbiased estimate would be a weighted average of the men's mean and the women's mean, with each sex weighted in proportion to its representation in the total population of 25-year-olds in the United States. Thus, in effect, each of the two women would be weighted approximately 5,000 times as much as each man. This clearly would result in an enormous standard error.

---

*Paper presented at a joint session of the American Psychological Association (Division 5) and Psychometric Society, in Chicago, 3 September 1975.

There are several ways our hypothetical investigator might cope with this problem but most of them create as many problems as they solve. One thing our investigator might do, if he felt that for purposes of his research he required an unbiased estimate of the mean and that he could tolerate a sizable random error, would be to decide that since there was inevitably a large standard error for his data on the female half of the population there would be little point in expending time, effort, and money in processing a large sample for the male half. He might therefore pick a random sample of two men from his 10,000, and use just those two to represent all the men in the United States. He would now have a four-case sample and would no longer have to weight each woman 5,000 times as much as each man in order to get an unbiased estimate. But this radical reduction of the male sample would do nothing to reduce the standard error; in fact it would increase it quite a bit.

But suppose that the investigator is willing to have a somewhat biased estimate in order to reduce the random error. How should he compute the biased estimate?

One way would be to use the estimate based on men only, and ignore the data for the women, on the grounds that those were the data responsible for almost all of the random error. This would have the disadvantage of producing a very large error due to bias.

Another possibility would be to get an unweighted mean of the 10,000 mean and 2 women. This would give an almost identical result to that obtained by omitting the women altogether since the men would have 5,000 times as large a representation as the women. This procedure would give as small a random error as would be possible with the available data, but the systematic error would be almost exactly as large as if only men were included.

3. An analogous real example (Project TALENT)

The hypothetical situation which has been described, in which one segment of the population is 5,000 times as well represented as another, equally important segment, is admittedly imaginary. But an analogous situation, one in which the relevant ratios are not quite so extreme, has actually occurred in educational research--in Project TALENT, to be specific. In Project TALENT a 400,000-case probability sample of students in grades 9-12 was tested in 1960 and we have been following them up by mail questionnaire ever since (one year, five years, and eleven years after their class graduated from high school). This means that almost every year since 1960 we have been sending out a questionnaire to everybody in our sample in a particular class--let's say about 100,000 sample members. Obviously not all of these 100,000 young men and young women reply; and equally obviously, those that do reply are not a random sample of the total group. In order to have the nonrespondents represented, we select a small systematic probability sample of them--let's say 4% of them--and make every effort to locate them and secure their answers to the questionnaire, usually be telephone interview. Because it is a systematic probability sample it is somewhat more precise than a random sample would be. (The systematic procedure consists in arranging the nonrespondents in identification-number order, which automatically arranges them by geographical region, by state within region, by city within state, by school, and in most schools, by classroom within school. To select a 4% sample, every

25th case would be chosen.) If we choose to regard the entire TALENT sample as constituting a finite population (which it is convenient to do in the present context*), we find that the population has been divided into two segments--(1) the respondents to the mailed questionnaire and (2) the nonrespondents. We can regard those who responded to the mailed questionnaire as a 100% sample of segment 1 and those in the 4% sample of nonrespondents as a 4% sample of segment 2.

We then have the problem of combining these two samples--the 100% sample and the 4% sample--to estimate the mean of the entire TALENT sample--i.e., the mean of the finite population. There are many ways this might be done:

> We can obtain an unbiased estimate by weighting every case in the 4% sample 25 times as much as it would have been weighted if everybody had responded to the mailed questionnaire. This estimate will have no systematic error but it may have a substantial random error. (Call this "Method a.")
>
> Or we can obtain a 4% sample of the respondent segment, to combine with the 4% of the nonrespondents. This will give us another unbiased estimate, with a larger random error but with considerable savings in processing costs. The percentage reduction in processing costs may be considerably greater than the percentage increase in random error. (Call this "Method b.")
>
> Or we can obtain an unweighted estimate--in which every case, whether in the 4% sample or in the respondent group, is weighted

---

*After determining the variance error of a statistic in terms of its deviation from the corresponding parameter for the finite population, it is quite easy to convert to variance error around the parameter for the hypothetical infinite population of which the "finite population" is conceived to be a random sample, by utilizing the additivity property of variance errors (the variance error of a sum is the sum of the variance errors)--provided, of course that one knows the variance error of the statistic for the "finite population."

equally. This will result in biased statistics, but they will have comparatively small random errors. (Call this "Method c.")

Or we can ignore the 4% sample altogether, and accept a very large systematic error, in order to eliminate random error. (Call this "Method d.")

Or we can reject all of these possibilities and look for a better way of combining the two sets of data--a way that will not only have a smaller random error than Methods a and b (the two methods of getting unbiased estimates of the mean) and a smaller systematic error than Methods c and d, but will have a smaller total error than any of them.

4. Statement of the problem in specific terms

Stated specifically, the problem with which this paper is concerned is how to combine data from two samples, representing two different segments of the population, to get an estimate of the mean for the total poulation, with the least-squares total error, where total error is defined as the sum of systematic and random errors.

The solution should be in terms which will make it applicable to either infinite or finite populations. And from the point of view of applicability to Project TALENT, the solution should cover the situation where rather than a simple random sample, the sample from one (or both) of the population segments is a systematic probability sample of some kind.

B. DERIVATION OF THE FORMULAS

1. Notation

The notation used in the algebraic development that follows is summarized in Table 1.

2. Derivation of general formulas (for the 2-group case)

Estimated mean: $\bar{X}_T = w_1\bar{X}_1 + w_2\bar{X}_2$ (1)

Sampling error of $\bar{X}_T$:

$$\sigma_{\bar{X}_T} = \sqrt{\sigma_{w_1\bar{X}_1}{}^2 + \sigma_{w_2\bar{X}_2}{}^2}$$

$$= \sqrt{w_1{}^2\sigma_{\bar{X}_1}{}^2 + w_2{}^2\sigma_{\bar{X}_2}{}^2} \qquad (2)$$

S.E. of mean of stratified sample from finite population:

$$\sigma_{\bar{X}_i} = \frac{s_{X-\bar{X}_{c_i}}}{\sqrt{n_i - c_i}} \cdot \sqrt{\frac{N_i - n_i}{N_i}} \qquad (3)$$

$$= \frac{\sigma_{X-\bar{X}_{c_i}}}{\sqrt{n_i}} \cdot \sqrt{\frac{N_i - n_i}{N_i - c_i}} \qquad (4)$$

Note that a simple random sample may be regarded as a special case of the stratified sample in which there is only one stratum or cell ($c_i = 1$).

Bias: $b_T = E(\bar{X}_T - \mu_T)$ (5)

$= E(\bar{X}_T) - \mu_T$

$= E(w_1\bar{X}_1 + w_2\bar{X}_2) - (\pi_1\mu_1 + \pi_2\mu_2)$

$= w_1E(\bar{X}_1) + w_2E(\bar{X}_2) - \pi_1\mu_1 - \pi_2\mu_2$

$= w_1\mu_1 + w_2\mu_2 - \pi_1\mu_1 - \pi_2\mu_2$

$= (w_1 - \pi_1)\mu_1 + (w_2 - \pi_2)\mu_2$

$= (w_1 - \pi_1)\mu_1 + [(1 - w_1) - (1 - \pi_1)]\mu_2$

$= (w_1 - \pi_1)(\mu_1 - \mu_2)$ (6)

$b_T{}^2 = (w_1 - \pi_1)^2(\mu_1 - \mu_2)^2$ (7)

Total error: y = squared sampling error + squared bias

$= \sigma_{\bar{X}_T}{}^2 + b_T{}^2$ (8)

$= (w_1{}^2\sigma_{\bar{X}_1}{}^2 + w_2{}^2\sigma_{\bar{X}_2}{}^2) + (w_1 - \pi_1)^2(\mu_1 - \mu_2)^2$

$= w_1{}^2\sigma_{\bar{X}_1}{}^2 + (1 - w_1)^2\sigma_{\bar{X}_2}{}^2 + (w_1 - \pi_1)^2(\mu_1 - \mu_2)^2$ (9)

Differentiating with respect to w_1:

$$y' = 2w_1\sigma_{\bar{X}_1}{}^2 - 2\,(1 - w_1)\sigma_{\bar{X}_2}{}^2 + 2\,(w_1 - \pi_1)(\mu_1 - \mu_2)^2 \quad (10)$$

Second derivative:

$$y'' = 2\,\sigma_{\bar{X}_1}{}^2 + 2\,\sigma_{\bar{X}_2}{}^2 + 2(\mu_1 - \mu_2)^2 > 0 \quad (11)$$

$\therefore$ y is a minimum when $y' = 0$

Setting $y' = 0$:

$$w_1 = \frac{\sigma_{\overline{X}_2}{}^2 + \pi_1(\mu_1 - \mu_2)^2}{\sigma_{\overline{X}_1}{}^2 + \sigma_{\overline{X}_2}{}^2 + (\mu_1 - \mu_2)^2} \tag{12}$$

$$= \frac{1 + \pi_1\left(\dfrac{\mu_1 - \mu_2}{\sigma_{\overline{X}_2}{}^2}\right)^2}{1 + \dfrac{\sigma_{\overline{X}_1}{}^2}{\sigma_{\overline{X}_2}{}^2} + \left(\dfrac{\mu_1 - \mu_2}{\sigma_{\overline{X}_2}}\right)^2} \tag{13}$$

Let $z = \dfrac{\mu_1 - \mu_2}{\sigma_{\overline{X}_2}}$

Substituting in (13):

$$w_1 = \frac{1 + \pi_1 z^2}{1 + \dfrac{\sigma_{\overline{X}_1}{}^2}{\sigma_{\overline{X}_2}{}^2} + z^2} \tag{14}$$

Substituting (4) in (14):

$$w = \frac{1 + \pi_1\left(\dfrac{\mu_1 - \mu_2}{\sigma_{X-\overline{X}_{c_2}}}\right)^2 \left(\dfrac{n_2(N_2 - c_2)}{N_2 - n_2}\right)}{1 + \left(\dfrac{\sigma_{X-\overline{X}_{c_1}}}{\sigma_{X-\overline{X}_{c_2}}}\right)^2 \dfrac{N_1 - n_1}{N_1 - c_1} \cdot \dfrac{N_2 - c_2}{N_2 - n_2} \cdot \dfrac{n_2}{n_1} + \left(\dfrac{\mu_1 - \mu_2}{\sigma_{X-\overline{X}_{c_2}}}\right)^2 \left(\dfrac{n_2(N_2 - c_2)}{N_2 - n_2}\right)} \tag{15}$$

Formulas 12 through 15 are general formulas for the value of w_1 that minimizes total error.

3. Special cases

a. No sampling error

In the special situation where $p_1 = p_2 = 1$ there is no random error (no sampling error). Therefore:

$$\sigma_{\overline{X}_1} = \sigma_{\overline{X}_2} = 0$$

Formula 12 therefore reduces to:

$$w_1 = \pi_1$$

b. No bias

In the special situation where there is no bias and where simple random sampling, rather than systematic sampling, was used, the following conditions would hold:

$$\sigma_{X-\overline{X}c_i} = \sigma_{X_i}$$

$$c_i = 1$$

$$\mu_1 = \mu_2$$

$$X_1 = X_2$$

Formula 15 would then reduce to:

$$w_1 = \frac{n_1}{n_1 + n_2 \dfrac{N_1 - n_1}{N_1 - 1} \quad \dfrac{N_2 - 1}{N_2 - n_2}}$$

$$= \frac{n_1}{n_1 + n_2 \dfrac{1 - \dfrac{n_1}{N_1}}{1 - \dfrac{n_2}{N_2}} \dfrac{1 - \dfrac{1}{N_2}}{1 - \dfrac{1}{N_1}}}$$

$$= \frac{n_1}{n_1 + n_2 \dfrac{1 - p_1}{1 - p_2} \dfrac{1 - \dfrac{1}{N_2}}{1 - \dfrac{1}{N_1}}} \qquad (16)$$

With the additional assumption that N_1 and N_2 are either very large or approximately equal, formula 16 becomes:

$$w_1 = \frac{n_1}{n_1 + n_2\left(\dfrac{1 - p_1}{1 - p_2}\right)} \qquad (17)$$

With the further assumption that p_1 and p_2 are approximately equal, formula 17 becomes, approximately:

$$w_1 = \frac{n_1}{n_1 + n_2} \qquad (18)$$

$$= \frac{p_1 N_1}{p_1 N_1 + p_2 N_2} = \frac{pN_1}{pN_1 + pN_2}$$

$$= \frac{N_1}{N_1 + N_2} \qquad (19)$$

c. Infinite populations

When $N_1 = N_2 = \infty$ formula 15 reduces to:

$$w_1 = \frac{n_1 + \pi_1 n_1 n_2 \left(\frac{\mu_1 - \mu_2}{\sigma_{X-\bar{X}_{c_2}}} \right)^2}{n_1 + n_2 \left(\frac{\sigma_{X-\bar{X}_{c_1}}}{\sigma_{X-\bar{X}_{c_2}}} \right)^2 + n_1 n_2 \left(\frac{\mu_1 - \mu_2}{\sigma_{X-\bar{X}_{c_2}}} \right)^2} \qquad (20)$$

In the event that $\sigma_{X-\bar{X}_{c_1}}$ can reasonably be assumed to equal $\sigma_{X-\bar{X}_{c_2}}$, the formula simplies further, to:

$$w_1 = \frac{n_1 + \pi_1 n_1 n_2 \left(\frac{\mu_1 - \mu_2}{\sigma_{X-\bar{X}_c}} \right)^2}{n_1 + n_2 + n_1 n_2 \left(\frac{\mu_1 - \mu_2}{\sigma_{X-\bar{X}_c}} \right)^2} \qquad (21)$$

If instead of systematic samples one has simple random samples, formula 21 further simplifies to:

$$w_1 = \frac{n_1 + \pi_1 n_1 n_2 z^2}{n_1 + n_2 + n_1 n_2 z^2} \qquad (22)$$

d. The Project TALENT application:

In the estimation of means for the entire Project TALENT population on the basis of those for whom questionnaire responses were available, if group 1 consists of the respondents to the mailed questionnaire and group 2 is nonrespondents, the following conditions apply:

$$\left.\begin{array}{l} N_1 < \infty \\ \\ N_2 < \infty \end{array}\right\} \text{Finite population}$$

$$n_1 = N_1 \quad \left.\begin{array}{r} (\text{Therefore } \sigma_{\overline{X}_1} = 0 \\ \\ \text{and} \quad \mu_1 = \overline{X}_1 \end{array}\right\} \begin{array}{l} \text{The sample for one} \\ \text{of the two populations} \\ \text{is the entire pop-} \\ \text{ulation.} \end{array}$$

$$n_2 < N_2$$

$$c_1 = 1 \quad (\text{Therefore } \sigma_{X-\overline{X}_{c_1}} = \sigma_{\overline{X}_1}.)$$

$c_2 = n_2$ (In other words the systematic probability sample of nonrespondents may be regarded as conceptually equivalent to a stratified random sample where each case in the sample is for a different stratum.)

In this situation, formula 15 reduces to:

$$w_1 = \frac{1 + \pi_1 n_2 \left(\dfrac{\overline{X}_1 - \mu_2}{\sigma_{X-\overline{X}_{c_2}}} \right)^2}{1 + n_2 \left(\dfrac{\overline{X}_1 - \mu_2}{\sigma_{X-\overline{X}_{c_2}}} \right)^2}$$

$$= \frac{\sigma_{X-\overline{X}_{c_2}}{}^2 + \pi_1 n_2 \, (\overline{X}_1 - \mu_2)^2}{\sigma_{X-\overline{X}_{c_2}}{}^2 + n_2 \, (\overline{X}_1 - \mu_2)^2} \qquad (23)$$

C. APPLICATION OF THE FORMULAS

Most of the formulas for w_1 require some values which are not known and cannot be computed precisely -- for instance μ_1, μ_2, $\sigma_{X-\overline{X}_{c_1}}$, and $\sigma_{X-\overline{X}_{c_2}}$. They therefore have to be estimated somehow. We may use the sample means ($\overline{X}_1$ and $\overline{X}_2$) as estimates of the population means (μ_1 and μ_2 respectively). The procedures used for estimating $\sigma_{X-\overline{X}_c}$ will depend on the nature of the systematic sampling and the nature of the available data.

To use formula 22 (the general formula for the Project TALENT application discussed above), we assume that $\mu_2 = \overline{X}_2$, but we also have to estimate, somehow, the value of $\sigma_{X-\overline{X}_{c_2}}$ when $c_2 = n_2$, in other words when there is just one case per cell.

This can be done by assuming that the larger c_2 is (in other words the closer it is to n_2), the lower the corresponding value of $\sigma_{X-\overline{X}_{c_2}}$. Utilizing this assumption, it is suggested that $\sigma_{X-\overline{X}_{c_2}}$ be computed for the following four values of d_2, where d_2 = cell density -- the number of cases per cell:

$d_2 = 5$

$d_2 = 4$

$d_2 = 3$

$d_2 = 2$

For instance for $d_2 = 3$ the estimate of the population standard deviation around the cell mean would be computed on the basis of all possible sets of three adjacent cases in the nonrespondent group.

The sets of three would of course overlap, since each case (except those at the ends) would be in exactly three triads.

After the four estimates of $\sigma_{X-\overline{X}_{c_2}}$ are obtained they would be extrapolated downward to get an estimate for $d_2 = 1$.

In applying formula 22 the researcher of course has to bear in mind that the values used for μ_2 and for $\sigma_{X-\overline{X}_{c_2}}$ are just estimates, not precise figures, and he should use them warily.

For instance suppose that $\overline{X}_2$ happened to be exactly equal to $\overline{X}_1$. If we set μ_2 equal to $\overline{X}_2$ under those circumstances formula 22 would simplify to:

$$w_1 = 1$$

This would happen irrespective of whether simple random sampling or systematic sampling had been used. The only essential conditions would be that:

$$n_1 = N_1$$

$$n_2 < N_2$$

$$\overline{X}_1 = \overline{X}_2$$

$$c_1 < N_1$$

D. RECOMMENDATIONS

1. Recommendations regarding use of the formulas

The procedure which has been developed is recommended for use when a combined mean is to be obtained from the means of two non-equivalent groups whenever it is important to minimize total error. This does not mean that the procedure should be used whenever a mean is to be obtained from combined groups.

In some cases it will not be feasible to obtain all of the necessary input data -- for instance it may not always be possible to obtain a reasonably good estimate of $\sigma_{\overline{X}-\overline{X}_c}$ when d = 1 (i.e., only one case per stratum). And in some cases $\overline{X}_1$ or $\overline{X}_2$ or both may have sufficiently large standard errors to deprive one of confidence in their use as replacements for μ_1 and μ_2 respectively. (This has already been pointed out in connection with formula 22.)

But even if all the necessary input data are readily available and accurate, the procedure presented here is not automatically the most appropriate procedure. In some circumstances it may be more appropriate to have an unbiased estimate than to minimize total error. It depends on the reason for obtaining the statistics, and the way they are to be used. In other words the ultimate decision rests with the researcher; he has to know what he is doing, and why.

2. Suggestions for further generalization of the procedure

Formulas have been presented for various applications in the two-group situation. A logical next step might be to generalize the procedure to three or more groups. It won't be as easy as it may sound -- but it would be useful.

Appendix F

TECHNICAL MEMORANDUM*

Statistical Procedure for Inferring Score Distribution of Unselected Group of 18-Year-Olds from Score Distributions of Unselected Group of 15-Year-Olds and Twelfth-Grade Students

By

Marion F. Shaycoft

AIR Project C-33
Contract AF 41(657) - 324

American Institutes for Research
Washington, D.C.
August 1963

AIR-C33-8/63-TM

---

* This is a very slightly re-edited and abridged version.

--MFS, December 1976

Statistical Procedure for Inferring Score Distribution of Unselected Group of 18-Year-Olds from Score Distributions of Unselected Group of 15-Year-Olds and Twelfth-Grade Students

By Marion F. Shaycoft

I. Introduction

Project TALENT data are available for probability samples of ninth-, tenth-, eleventh-, and twelfth-grade students, and for a probability sample of 15-year-olds. This technical memorandum has as its function to describe a method of inferring the test score distribution of the 18-year-old population on the basis of the corresponding distributions for 15-year-olds and twelfth-grade students, and to present the mathematical derivation of the formulas used.

This method was applied as a part of a project carried out for the Air Force by the American Institute for Research*. In that project norms for an unselected 18-year-old male population were estimated from distributions with cases weighted to represent an unselected sample of 15-year-old boys and an unselected sample of 18-year-olds.

II. Underlying Assumption

Basically the procedure consists in assuming that most twelfth-grade boys are about 18 years old** and that those boys who have failed to reach Grade 12 by the time they are 18 or have become dropouts by then represent the bottom portion of a hypothetical normally distributed underlying variable representing "tendency to make normal progress through school". The truncated distribution represents what is left after the bottom tail of the distribution, containing the dropouts and the 18-year-olds who are below Grade 12, has been eliminated. It must be borne in mind that the distribution we are talking about here is not a score distribution but the continuous distribution on a hypothetical construct variable which can be dichotomized to have perfect negative correlation with the real variable representing "dropout-or-academic-retardation" behavior.

---

*Dailey, J.T., Shaycoft, Marion F., & Orr, D.B. Calibration of Air Force Selection Tests to Project Talent Norms. Washington, D.C., May 1962. (Final report on Air Force Contract No. AF 41(657)-324.)

**Actually more twelfth-graders in Project TALENT had not yet reached their 18th birthday than had passed it by the time they were tested (Spring 1960 mostly in March and April). It is assumed, however, that the growth curve representing increase in score with age flattens out sufficiently by age 17 or 18 for most of the tests involved, that the assumption that 17-year-olds are about the same as 18-year-olds does not have a serious effect on the results of applying the procedure described in this memorandum.

The other basic assumption is that the test score distribution for the complete age group (e.g., 18-year-olds) is approximately normal.*

III. Notation and Terminology

The following rules apply to most of the notation used in this derivation:

1. Capital letters represent statistics based on an unselected sample of a complete age group.

2. Small letters are used for statistics based on a curtailed sample (e.g., a grade group, which, for purposes of this derivation, is assumed to coincide essentially with an age group except for elimination, largely by dropout, of the tail at the low end on the variable representing tendency to make normal progress through school.)

Variables

X_1 = hypothetical normally distributed variable representing tendency to make at least normal progress through school. (This variable is assumed to be a monotonically increasing function of last grade completed.)

z_1 = corresponding standard score

X_2 = raw score on test

z_2 = corresponding standard score

---

* This assumption is necessary since biserial correlation coefficients are used in the procedure. Inherent in the formula for biserial r is the assumption that the continuous variable is normally distributed.

Selected group

The "selected group" consists of cases in the right-hand tail, (i.e., the non-dropout tail) when the total group has been dichotomized on variable X_1.

p = proportion of cases in selected group

z' = corresponding standard score (i.e., standard score corresponding to "100q" percentile where $q = 1-p$)

y' = corresponding normal ordinate

Means

$\bar{X}_i$ = raw mean on Variable X_i for total (unselected) group

$\bar{X}_{p_i}$ = corresponding mean on Variable X_i for selected group

$\bar{z}_{p_1}$ = mean standard score on Variable 1 for selected group

$\bar{z}_{p_2}$ = mean standard score on Variable 2 for selected group (i.e., for group selected on basis of Variable 1)

Standard deviations

S_i = standard deviation on Variable i for total (unselected) group

s_i = standard deviation on Variable i for selected group

Correlations

R_{12} = correlation between X_1 and X_2 in unselected group

r_{12} = corresponding correlation in selected group

IV. Derivation

From formulas for mean and standard deviation of the tail of a normal distribution:

$$\bar{z}_{p_1} = \frac{y'}{p} \qquad (1)^{*}$$

$$s_{z_1} = \sqrt{1 + \frac{y'z'}{p} - \left(\frac{y'}{p}\right)^2} \qquad (2)^{**}$$

Formulas for correction of r_{12} for curtailment due to cut on Variable 1:

$$R_{12} = \frac{r_{12}\,\frac{S_1}{s_1}}{\sqrt{1 - r_{12}^2 + r_{12}^2\,\frac{S_1^2}{s_1^2}}} \qquad (3)^{***}$$

$$R_{12} = \sqrt{1 - \frac{s_2^2}{S_2^2}\left(1 - r_{12}^2\right)} \qquad (4)^{****}$$

---

* Kelley (3) Formula 8:26

** The derivation of this formula is shown starting on page F-14.

*** Thorndike (4) p. 173 Formula 6

**** Thorndike (4) p. 173 Formula 5

From formula #3 above:

$$r_{12} = \frac{R_{12}\,\frac{s_1}{S_1}}{\sqrt{1 - R^2_{12} + R^2_{12}\,\frac{s^2_1}{S^2_1}}}$$

$S_{z_1} = 1$ (from definition of standard score)

$$\therefore\ r_{12} = \frac{R_{12}\,s_{z_1}}{\sqrt{1 - R^2_{12} + R^2_{12}\,s^2_{z_1}}} \qquad (5)$$

From formula #4 above:

$$\frac{s^2_2}{S^2_2} = \frac{1 - R^2_{12}}{1 - r^2_{12}}$$

$$\therefore \quad s_2 = S_2 \sqrt{\frac{1 - R^2_{12}}{1 - r^2_{12}}}$$

$S_{z_2} = 1$ (from definition of a standard score)

$$\therefore \quad s_{z_2} = \sqrt{\frac{1 - R^2_{12}}{1 - r^2_{12}}} \qquad (6)$$

$$\frac{S_{x_2}}{s_{x_2}} = \frac{S_{z_2}}{s_{z_2}} = \frac{1}{s_{z_2}}$$

$$\therefore \quad S_{x_2} = \frac{s_{x_2}}{s_{z_2}} \qquad (7)$$

$$\text{Biserial } R_{12} = \frac{\bar{X}_{p_2} - \bar{X}_2}{S_{x_2}} \cdot \frac{p}{y'}$$

$$= \bar{z}_{p_2} \cdot \frac{p}{y'}$$

$$\bar{z}_{p_2} = \frac{y'}{p} R_{12\,(\text{bis})} \tag{8}$$

$$\bar{z}_{p_2} = \frac{\bar{X}_{p_2} - \bar{X}_2}{S_{x_2}}$$

$$\bar{X}_2 = \bar{X}_{p_2} - \bar{z}_{p_2} S_{x_2} \tag{9}$$

Combining formulas (7), (6), (5) :

$$\begin{aligned}
S_{x_2} &= \frac{s_{x_2}}{s_{z_2}} \\
&= \frac{s_{x_2}\sqrt{1 - r^2_{12}}}{\sqrt{1 - R^2_{12}}} \\
&= \frac{s_{x_2}}{\sqrt{1 - R^2_{12}}} \cdot \frac{\sqrt{1 - R^2_{12} + R^2_{12}\, s^2_{z_1} - R^2_{12}\, s^2_{z_1}}}{\sqrt{1 - R^2_{12} + R^2_{12}\, s^2_{z_1}}} \\
&= \frac{s_{x_2}}{\sqrt{1 - R^2_{12}}} \cdot \frac{\sqrt{1 - R^2_{12}}}{\sqrt{1 - R^2_{12} + R^2_{12}\, s^2_{z_1}}}
\end{aligned}$$

$$S_{x_2} = \frac{s_{x_2}}{\sqrt{1 - R_{12}^2 + R_{12}^2 s_{z_1}^2}} \qquad (10)$$

where $$s_{z_1}^2 = 1 + \frac{\underline{y'z'}}{p} - \left(\frac{\underline{y'}}{p}\right)^2 \qquad (2)$$

and R_{12} = biserial correlation coefficient (estimated).

V. Application of procedure to Airman Qualifying Examination (AQE) Aptitude Indexes

In the project previously referred to that was carried out for the Air Force, the procedure described above was used to develop norms for 18-year-old males on the four AQE Aptitude Indexes.

For this purpose estimates were obtained of the correlation between last grade completed and each of the Aptitude Indexes (for Air Force recruits). These estimates are quite rough, since they are based on data obtained in 1948*, as modified by subjective judgment concerning the probable direction and magnitude of the changes that have occurred since then. While the values used for biserial r are thus admittedly subject to considerable error, there is evidence, fortunately, that such error is not likely to have a large effect on the resultant estimates of the means and standard deviations for the unselected population, since these estimates are apparently relatively insensitive to the precise value used for the biserial correlation. (Exploratory application of the procedure to the TALENT Reading Comprehension Test, for instance, has revealed that changing the estimate of the biserial correlation from .50 to .40 increases the estimated mean by only 2.2 percent and decreases the estimated standard deviation by only 2.7 percent.

Table 1 shows the statistics on which the estimates of the mean and standard deviation for the TALENT composites corresponding to the four Aptitude Indexes are based, the intermediate computational steps, and the resultant estimates.

After these estimates were obtained the next step was handled graphically. Percentiles corresponding to each TALENT composite raw score were plotted on normal percentile graph paper, for 15-year-old boys. On these graphs the percentile was the horizontal scale and the raw score was the vertical scale. Some of these variables turned out not to have perfectly normal distributions. In other words the graphs in general were not perfectly straight lines. The magnitude and direction of the departure from normality was indicated by plotting three additional points on the same graph: the raw score mean, which was plotted at the 50th percentile point, the "mean-plus-standard-deviation" value, which was plotted at the 84.1 percentile, and the "mean-minus-standard-deviation", which was plotted at the 15.9 percentile. These three points were not in general on the percentile curve, of course.

---

* Dailey, J.T. The development of a comprehensive aptitude battery for Air Force Technical Specialties. Unpublished doctoral dissertation, University of Texas, 1949.

The corresponding three points were then plotted for the 18-year-olds, using the estimated values of the means and standard deviations, the computation of which is shown in Table 1. A curve was then drawn for the 18-year-olds which appeared visually to have the same characteristics as the curve for the 15-year-olds except, of course, for the different means and standard deviations. By this is meant that the percentile curve for the 18-year-olds was drawn to have about the same shape as the curve for the 15-year-olds and to have about the same placement with respect to the three special points (mean, mean-plus-standard-deviation, and mean-minus-standard-deviation). A very minor deviation from the "similar shape" principle occurred on a couple of the graphs above the 99.9 percentile point, where it was necessary to induce a slight negative acceleration* in the curve to force it to approach the maximum possible raw score asymptotically rather than exceed it. (This induced bend is a matter of theoretical interest only, not a practical consideration, since points above the 99.9 percentile affect hardly anyone, and do not alter the 99th percentile.)

---

*"Negative second derivative", to use calculus terms.

Table 1

[illegible]mputation of estimated means and standard deviations for total 18-year-old male population on four TALENT composites corresponding to AQE aptitude indexes

X_2	Source Data			Intermediate Computation			Final Estimates (for total group of 18-year-old boys)	
		Grade 12 boys						
TALENT composite for:	R_{12}	$\bar{X}_{p_2}$	s_{x_2}	$1-R^2_{12}+R^2_{12}s^2_{z_1}$	$\sqrt{1-R^2_{12}+R^2_{12}s^2_{z_1}}$	$\bar{z}_{p_2}$	$\bar{X}_2$	S_{x_2}
AQE Admin. Apt. Index	.50	176.13	40.52	.8644	.9297	.28272	163.81	43.58
AQE Mech. Apt. Index	.50	97.73	23.04	.8644	.9297	.28272	90.72	24.78
AQE Gen. Apt. Index	.65	127.63	36.91	.7709	.8780	.36754	112.18	42.04
AQE Elec. Apt. Index	.60	92.36	28.14	.8048	.8971	.33926	81.72	31.37
	$p = .653$			$\bar{z}_{p_1} = y'/p = .56544$ $z' = -.3934$ $s^2_{z_1} = .4578$				

Notation:

X_1 = last grade completed

X_2 = TALENT composite for AQE aptitude index

p = Proportion of Grade 5 entrants who reach Grade 12

Assumption: For present purposes it is assumed that the total 18-year-old male population has approximately the same distribution (on the four TALENT composites) as the total Grade 12 male population would have if there were no dropouts

Source of data:

R_{12} is the assumed approximate value of the biserial correlation. The assumed values are estimates based on data provided by Dr. John T. Dailey (1).

$\bar{X}_{p_2}$ and s_{x_2} are from (2) Table 10.

p is U.S. Office of Education data (published in *Information Please Almanac,* 1962).

Computational procedure:

$\bar{z}_{p_1}$ and z' are from a normal distribution table showing abscissas, ordinates, areas, etc.

$s^2_{z_1}$ is from formula 2, which is derived in the Appendix, where it appears as Formula 20.

$\bar{z}_{p_2}$ is from formula 8

$\bar{X}_2$ is from formula 9

S_{x_2} is from formula 10

REFERENCES

1. Dailey, J.T. The development of a comprehensive aptitude battery for Air Force Technical Specialties. Unpublished doctoral dissertation, University of Texas, 1949.
2. Dailey, J.T., Shaycoft, Marion F., & Orr, D.B. Calibration of Air Force Selection Tests to Project Talent Norms. Washington, D.C., May 1962. (Final report on Air Force Contract No. AF 41(657)-324.)
3. Kelley, T.L. Fundamentals of Statistics. Harvard University Press, Cambridge, Massachusetts, 1947.
4. Thorndike, R.L. Personnel Selection - Test and Measurement Techniques. John Wiley & Sons, Inc., New York, 1949.

I. Introduction

The purpose of this appendix is to present a derivation of the formula indicated by the title above. Incidental by-products of this derivation are derivations of formulas for the standard deviation of a non-tail segment of a normal distribution, the mean of a non-tail segment, and the mean of a tail.

II. Basic formulas, definitions, and notation

A. Equation for unit normal distribution

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{1}$$

B. Notation

p_i = proportion of cases with $z>z_i$
(See footnote *)

p_j = proportion of cases with $z>z_j$

p_{ij} = proportion of cases with $z_j>z>z_i$

y_i = normal ordinate corresponding to z_i
(See footnote *)

$\bar{z}_i$ = Mean of cases with $z>z_i$
(See footnote *)

$\bar{z}_{ij}$ = Mean of cases with $z_j>z>z_i$

$s^2_{z_i}$ = variance of cases with $z>z_i$

$s^2_{z_{ij}}$ = variance of cases with $z_j>z>z_i$

C. Area of segment ij

$$p_{ij} = p_i - p_j = \int_{z_i}^{z_j} y \, dz \tag{2}$$

---

* In the main body of the report the value $\bar{z}_i$ is called $\bar{z}_{p_1}$; p_i is called p; z_i is called z´; and y_i is called y´.

D. Area of tail

$$p_i = \int_{z_i}^{\infty} y \, dz \qquad (3)$$

III. Derivation

A. Determining mean of cases in segment ij

$$\bar{z}_j = \frac{1}{p_{ij}} \int_{z_i}^{z_j} yz \, dz = - \frac{y}{p_{ij}} \Bigg]_{z_i}^{z_j}$$

$$= \frac{y_i - y_j}{p_{ij}} \qquad (4)$$

$$= \frac{y_i - y_j}{p_i - p_j} \qquad (5)$$

B. Determining mean of cases in tail, z_i to ∞

$$z_j = \infty$$

$$y_j = 0$$

$$p_j = 0$$

$$\bar{z}_i = \frac{y_i - 0}{p_i - 0} = \frac{y_i}{p_i} \qquad (6)$$

C. Determining the variance of segment ij

$$s^2_{z_{ij}} = \frac{1}{P_{ij}} \int_{z_i}^{z_j} yz^2 dz - \left[\frac{1}{P_{ij}} \int_{z_i}^{z_j} yz\ dz \right]^2$$

$$= \frac{1}{P_{ij}} \int_{z_i}^{z_j} yz^2 dz - \left[\frac{y_i - y_j}{P_i - P_j} \right]^2 \tag{7}$$

Let yz dz = du

Then u = -y

$$\int yz^2 dz = \int z\ du = uz - \int u\ dz$$

$$= -yz + \int y\ dz$$

$$\therefore \int_{z_i}^{z_j} yz^2 dz = -yz \Big]_{z_i}^{z_j} + \int_{z_i}^{z_j} y\ dz$$

$$= -yz \Big]_{z_i}^{z_j} + P_{ij}$$

$$= \left[y_i z_i - y_j z_j \right] + \left[P_i - P_j \right] \tag{8}$$

Substituting (8) in (7)

$$s^2_{z_{ij}} = \frac{y_i z_i - y_j z_j}{p_{ij}} + \frac{p_i - p_j}{p_{ij}} - \left(\frac{y_i - y_j}{p_i - p_j}\right)^2$$

$$= 1 + \frac{y_i z_i - y_j z_j}{p_{ij}} - \left(\frac{y_i - y_j}{p_{ij}}\right)^2 \qquad (9)$$

$$= 1 + \frac{y_i z_i - y_j z_j}{p_{ij}} - \bar{z}^2_{ij} \qquad (10)$$

D. Determining variance of tail, z_i to ∞:

$z_j = \infty$

$y_j = 0$

$p_j = 0$

These values cannot be substituted directly in (9), since one of the terms in that equation, $y_j z_j$, would become an indeterminate quantity, $0 \cdot \infty$*.

---

* The standard method of evaluating an indeterminate of this sort, by converting to a 0/0 form and differentiating numerator and denominator separately, does not work in this instance, since the outcome is another indeterminate, $y_j z_j^3$, as shown below:

$$\lim_{z \to \infty} yz = \lim_{z \to \infty} \frac{y}{z^{-1}} = \lim_{z \to \infty} \frac{dy}{dz} \Big/ \frac{d(z^{-1})}{dz}$$

$$= \lim_{z \to \infty} \frac{-yz}{-z^{-2}} = \lim_{z \to \infty} yz^3 \longrightarrow 0 \cdot \infty$$

When $z_i = -\infty$ and $z_j = \infty$, (11)

$$s^2_{z_{ij}} = 1,\ \bar{z}_{ij} = 0,\ \text{and}\ p_{ij} = 1 \qquad (12)$$

by the definition of a unit normal distribution,

and equation (10) becomes:

$$\lim_{\substack{z_i \to -\infty \\ z_j \to \infty}} s^2_{z_{ij}} = 1 + \frac{\lim\limits_{z_i \to -\infty} y_i z_i - \lim\limits_{z_j \to \infty} y_j z_j}{p_{ij}} - \bar{z}^2_{ij} \qquad (13)$$

Substituting (12) in (13):

$$1 = 1 + \lim_{z_i \to -\infty} y_i z_i - \lim_{z_j \to \infty} y_j z_j$$

$$\therefore \lim_{z_i \to -\infty} y_i z_i = \lim_{z_j \to \infty} y_j z_j \qquad (14)$$

$$\text{But } \lim_{z_i \to -\infty} z_i = -\lim_{z_j \to \infty} z_j = \lim_{z_j \to \infty} (-z_j) \qquad (15)$$

And since the unit normal distribution is symmetrical,

$$\lim_{z_i \to -\infty} y_i = \lim_{z_j \to \infty} y_j \qquad (16)$$

Substituting (15) and (16) in (14):

$$\lim_{z_i \to -\infty} y_i z_i = \lim_{z_j \to \infty} y_j(-z_j)$$

$$= -\lim_{z_j \to \infty} y_j z_j \qquad (17)$$

From (14) and (17):

$$\lim_{z_j \to \infty} y_j z_j = -\lim_{z_j \to \infty} y_j z_j$$

$$\therefore \lim_{z_j \to \infty} y_j z_j = 0 \qquad (18)$$

From (10):

$$s^2_{z_i} = \lim_{z_j \to \infty} s^2_{z_{ij}}$$

$$= 1 + \frac{y_i z_i - \lim_{z_j \to \infty} y_j z_j}{p_i} - \left(\frac{y_i}{p_i}\right)^2 \qquad (19)$$

Substituting (18) in (19):

$$s^2_{z_i} = 1 + \frac{y_i z_i}{p_i} - \left(\frac{y_i}{p_i}\right)^2 \qquad (20)$$

A New Multivariate Index for Use in Educational Planning*

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

The late Phillip Rulon once wrote a delightful little fable about a Hexagon who didn't know whether it should become a Circle or a Square - because it had corners, like a Square, but was round like a Circle. The Hexagon was given conflicting advice by a Relentless Psychologist who told it it would be a superior Circle, and an Inveterate Statistician who told it it was really a Square, not a Circle at all. (Rulon wrote his article before the term "Square" had the connotation it has today.) The Inveterate Statistician used Discriminant Analysis to come to his conclusion, while the Relentless Psychologist used Stanines for the purpose. Rulon ended his fable with the following "moral":

> "The Multiple Correlation Technique applied successively to different groups yields information not given by the Discriminant Function applied to all the groups,
>
> and vice versa."

The moral is perfectly sound, but unfortunately the resulting conflicting advice left the poor little Hexagon in a state of perplexity, not knowing whether to get two of its corners rubbed off and become a Square or get all of its corners rubbed off and become a Circle.

What might have been added to the moral is that the multiple correlation approach and the multiple discriminant function approach, even in combination, don't yield all the relevant information available.

The term "multiple correlation approach", in this context, refers to the correlation of a set of predictors against a continuous criterion that distinguishes among members of the group (in terms of quality of performance or whatever other characteristic seems relevant). The term "multiple discriminant function approach" as used here refers to the situation where the total sample has been divided into several groups and the sole criterion is membership in a group. There is no continuous criterion available to distinguish among members of the same group. The discriminant functions are a set of linear composites of the original predictor variables, and considered in combination these discriminant functions locate the group centroids in a space which includes all dimensions in which they differ, and maximizes the between-groups variance relative to the within-groups variance, in that space.

---

*Paper presented at the American Psychological Association Convention, in Washington on September 2, 1969, as part of a symposium on "New Approaches to Educational Planning".

Actually in addition to these two approaches there is a third commonly used one, which is really a cross between the first two. This is the approach where even though there are more than two groups, instead of performing one multiple discriminant analysis that takes all groups into account simultaneously, a separate discriminant analysis is performed for each group, to discriminate between that group and all other groups combined. It is well known, of course, that this results in a single discriminant function for each group, and that this discriminant function is identical (except for its standard deviation) with the multiple regression equation against the dichotomous criterion of group membership. In other words it is equivalent to the results of a conventional multiple correlation approach where point biserial correlations are used for the predictor-criterion correlation coefficients. I shall therefore refer to this as the "multiple point biserial R" approach, by way of shorthand.

In addition to these three techniques (multiple correlation, multiple discriminant analysis, and multiple point biserial R), I am proposing a fourth technique, called "propinquity analysis", which superficially has some of the features of each of the other approaches but is actually quite different from any of them.

Before I go into the details of this procedure, let me tell you what Rulon's hexagon should become according to the results of propinquity analysis. At this meeting of psychologists I hate to disagree with the "Relentless Psychologist", but the fact is that the Hexagon resembles a Square much more closely than it resembles a Circle. Its propinquile for being a Square is about 85, while its propinquile for being a Circle is less than 10.

A propinquile, I should explain, is the percentile corresponding to the propinquity index.

THE INADEQUACIES OF CONVENTIONAL METHODS

Perhaps the simplest way to describe what the propinquity index is and does is to start with a discussion of the problems we have to bypass - the problems and disadvantages inherent in any guidance system that depends solely or heavily on discriminant functions. I want to stress that what we are talking about in this connection is the guidance problem, not the personnel classification problem. Most of the difficulties and disadvantages in connection with multiple discriminant functions are of little or no relevance in connection with personnel classification, where there are specified numbers of vacancies in various occupational categories, with the total number of vacancies in all categories combined exactly matching the number of people to be assigned to the vacancies. The multiple discriminant function approach of course has many other appropriate uses, besides personnel classification. But here are some of the disadvantages of using it as the basis for guidance:

1. If more than one discriminant function is involved, which is

almost certain to be the case if there are at least two variables and at least three groups, generally each of these composite variables except the first is bipolar, having some positive and some negative correlations with the original variables. Moreover the set of "regression weights" applied to the original variables to compute any discriminant function beyond the first generally includes both positive and negative values of substantial magnitude. All of this is likely to make discriminant functions beyond the first very obscure in meaning - if not downright uninterpretable - at least in the realm of cognitive variables, where the concept of a truly bipolar factor seems inherently somewhat contradictory and therefore not very useful. Discriminant functions are admittedly artifacts, and there is little reason for using artifacts that are not basically helpful, merely because they happen to have some interesting statistical properties. The fact that discriminant functions maximize a certain kind of variance ratio when computed from a specified set of variables, with respect to a specified set of groups, doesn't automatically endow all discriminant functions with intrinsic importance or make them good tools for guidance.

2. But if bipolar discriminant functions are hard to interpret and give names to, how much harder it must be to try to explain them to someone else - someone vitally concerned with what they really mean. For instance it seems to me that it would be awfully difficult to explain to a boy who has his heart set on becoming a lawyer that he really shouldn't because his score is too low on a composite composed of (to take a strictly hypothetical example) two parts "Arithmetic Reasoning" plus one part "Memory for Words" minus one part "Spelling" plus two parts "English Usage" minus three parts "Visualization in Three Dimensions" ..... etc.

3. And what helpful suggestions can one give the would-be lawyer? That he should acquire the habit of misspelling, since the lower his Spelling score is the less will be subtracted? Or merely that the next time he takes such a battery he should deliberately give incorrect answers on those tests he knows to be negatively weighted on the discriminant function on which he wants to improve his score?

4. To get an idea of some even more serious disadvantages, look at Figure 1, on page 19 (in the Appendix). In this diagram the two axes, X_1 and X_2, represent scores on two uncorrelated variables with equal standard deviations. The large circle represents the total group, T. The center of the circle is the centroid of the group and the radius equals 2σ so that almost everyone in the group is within the circle. Likewise the two smaller solid circles represent subgroups, A and B. The total group, T, consist of three subgroups -- A, B, and an amorphous group containing everybody else. Since the centroids of A, B, and T are collinear, the centroid of the amorphous residual group, T-A-B, is necessarily also collinear with the A and B centroids; therefore in a discriminant analysis of Groups, A, B, and the residual group there would be only one discriminant function, which would be a rescaling of Variable X_1. The Variable X_2 data would be lost, although the fact that there is a marked restriction of Group A's range with respect to Variable X_2 suggests that that variable has some importance in determining a person's suitability for Group A. (To a lesser extent the same thing is true of Group B.) Individual I_1, who seems seriously deficient on whatever quality is measured by X_2, would be diagnosed as fitting into Group A optimally, if the discriminant analysis of Groups A, B, and the residual group were the sole basis of the decision. Individual I_2, who is apparently very much overqualified on X_2, would likewise be assigned to Group A. Thus discriminant analysis may produce a set of composites which omits one or more important dimensions.

5. Now let's look at the two dotted circles, C and D. Suppose that these are two additional groups, separated out from the residual group, and that a five-group multiple discriminant analysis is now done, on Groups A, B, C, D, and what remains of T. Because Groups C and D have entered the picture a second discriminant function, which is a rescaling of X_2, appears. It then becomes evident that cases I_1 and I_2 would be misfits in Group A despite their fine fit on the first discriminant function, because they are so far off on the second. But it seems unreasonable that an evaluation of whether one is well suited for a particular category should depend on whether certain other groups - ones that are definitely inappropriate - have been separated out.

The propinquity index and its close relative the propinquile were designed to be free of all these problems inherent in the multiple discriminant analysis approach. But there are also problems in connection with the multiple correlation and multiple point biserial R, and the propinquity statistics were designed to avoid those problems, too. Inherent in the multiple point biserial R approach (and also in the multiple correlation approach) is the assumption that the more one possesses of an ability or other trait that has a positive beta weight, the more suited one is for membership in the occupational group (or whatever other kind of group is involved), and that the more one possesses of an ability or trait that has a negative beta weight, the less suited one is for group membership.

Consider what this means in terms of an occupation which requires about average intelligence. If the incumbents in this group happen to have an average IQ of 98 and if the range of IQ's for this group happens to be between 90 and 107 (which is quite a restricted range), the beta weight is likely to be a small negative number, and someone with an IQ of 70 will appear more qualified than someone with an IQ of 100.

Now suppose that instead of a mean IQ of 98 the incumbents happen to have a mean IQ of 102, but the same restricted range. The beta weight for this situation is likely to be a small positive number, and someone with an IQ of 150 will appear to be better qualified than someone with an IQ of 101, when actually the person with an IQ of 150 would be grossly overqualified. And in view of the resultant waste of talent, overqualification for an occupation is about as undesirable as gross underqualification - at least it is in the square world of the non-hippy non-dropout-from society.

Now let's look at the third possibility - that the mean IQ for the group is very close to 100, still with a very restricted range, and that the beta weight happens to be 0. In this situation neither the person with an IQ of 150 nor the one with an IQ of 70 would be disqualified, although the restriction of range suggests that both should be.

The propinquity statistics take care of all these problems.

THE NATURE AND CHARACTERISTICS OF PROPINQUITY INDEXES

Briefly, an individual's propinquity index with respect to a given occupational group is his geometric distance* in n-dimensional space from the group centroid, where each dimension is the standard score on one of the original variables (standardized in terms of the group in question) weighted by a value representing, at least approximately, the relevance of the corresponding variable in identifying group members. A minus sign is attached to the distance, so that a propinquity index of 0, indicating that the individual's scores on relevant variables coincide with the group centroid, is the maximum value of the index.

---

*This represents a very slight change from the original definition, which was used in the oral presentation at APA. According to the original definition, the propinquity index was the square of the distance. The reason for the change was to improve some statistical characteristics of the index.

The greater the distance, the larger the absolute value of the index. The purpose of the minus sign, therefore, is to orient the propinquity index properly, so that when it is used in a correlation matrix there won't be inconvenient negative correlations. The higher the algebraic value of the index, the closer the individual is to the centroid. (Hence the term "propinquity index".)

Formula 12 (or 13) in the appendix* represents the squared distance in regard to a single variable. Formula 14 gives the propinquity index, δ, with w as the weight representing the relevance of a particular variable.

Formulas 22-28 are seven different formulas giving different results, that might be used for determining the w's. All these formulas have the desired characteristic of giving a weight of 0 for irrelevant variables and a positive weight for relevant variables. There are no negative weights. The indicator of relevance for a variable is a function of the ratio of the standard deviation within the group to the standard deviation of the total sample (all groups combined).

Empirical exploration is necessary in order to decide which of the seven formulas for w_{ij} is the one that should be used. It might occur to some of you that instead of using any of them we should just use multiple regression weights (against the dichotomous criterion of group membership). This is a possibility, but there are several factors militating against it. In the first place there is the bouncing beta phenomenon, which is likely to be a problem when the number of predictors is large - even if there is a very large number of cases. And the fact that the predictors are not merely linear functions of the variable, but involve squared terms, might make the betas bounce even more. Another disadvantage of regression weights lies in the very practical consideration that even with a high-powered computer, the computation of the betas for these squared terms is an extraordinarily complex, time-consuming, and expensive operation when the number of cases, number of variables, and number of groups are all comparatively large (as is true in the case of the Project TALENT data, to which this technique is being applied).

---

*All formulas in this paper are in the Appendix, which also contains a section defining all the notation used.

In view of the practical and theoretical objections to the use of multiple regression weights in this context, the best solution seems to be to see which of the seven systems of weights (Formulas 22-28) gives the weighted composites that have the highest point biserial correlation with the dichotomous group membership criterion.

All seven of the possible formulas for weights use some function of $\sigma_{z_{ij}}$ as the indicator of relevance of the variable. Since $\sigma_{z_{ij}}$ is the ratio of the standard deviation within the group to the total standard deviation (Formula 15) there is an assumption that a relevant variable will have a somewhat restricted range. and the more relevant the variable, the more restricted the range. This assumption may not be sound when some of the variables have skewed distributions. For such variables the ratio of group σ to total σ may be erratic, resulting in peculiar weights (w). Therefore it is desirable that all variables used in a propinquity analyses have normal distributions. If there is no reason to believe that the distributions are at least approximately normal they should be normalized.

Because the variables used in computing the propinquity index are raw scores rather than principal components, discriminant functions, or some other kind of uncorrelated composites, propinquity indexes may lack some of the mathematical precision and invariance of statistics computed in a geometric space where orthogonal axes correspond to uncorrelated variables. But this seems a small price to pay for the twin advantages of interpretability and ready explainability. Of course principal components or discriminant functions could be used as the initial scores, but it isn't recommended.

PROPINQUILES

In the univariate case the distribution of δ resembles the left half of a normal distribution, assuming the original variable is normally distributed. But when the number of variables entering into the propinquity index is greater than one, neither the distribution of δ nor the distribution of δ^2 is normal, nor are these distributions any portion of a normal distribution. Furthermore the basic shape of the distribution varies with the number of variables involved. Propinquity indexes based on three variables have an entirely different distribution from those based on two, and so forth.

And even when propinquity indexes for a set of groups are all based on the same battery they still may not be directly comparable, because though all indexes technically are based on the same number of variables, some variables may have zero weights for certain groups so that the number of dimensions is in effect different for different groups. Since propinquity indexes for different categories therefore are not generally directly comparable, they have to be converted to some uniform scale in order to be compared - and percentiles serve this purpose effectively and directly. These percentiles are the propinquiles.

THE ROLE OF PROPINQUITY ANALYSIS

Propinquity analysis is not in any sense of the word a replacement for the multiple correlation approach, nor for the multiple point biserial R approach. Rather, propinquity analysis should serve as an adjunct to both these procedures. In situations where the group membership dichotomous criterion is available, not only can the two procedures, propinquity analysis and multiple point biserial R, be used independently, but also the multiple point biserial R approach can be modified to incorporate the propinquity index as an additional predictor variable. This couldn't be done, of course, if the propinquity index were merely some sort of linear function of the n initial predictors, but it isn't, since squared terms are also involved. The propinquity index thus is not linearly dependent on the n variables which enter into it.

A few special considerations should be pointed out in regard to the use of the propinquity index as an additional predictor:

1. In the first place, the weights for the U components of the propinquity index should probably be determined on the basis of a different sample from the one used for the correlation matrix, in order to avoid capitalizing on chance in the point biserial correlation of the propinquity index with the dichotomous criterion, and to prevent this correlation from being spuriously high. The same thing applies if biserial correlations rather than point biserials are used, although it is not advisable to do this unless all the predictive variables, including the propinquity index, either have been normalized or are known to be normally distributed. Since two separate (but parallel) samples should be used on the combination procedure, a third parallel sample should also have been set aside initially for cross-validation.*

2. To use this combination procedure, a very large number of cases has to be available, not only because the initial sample has to be split into three separate samples in order to carry out the complete procedure properly, including cross-validation, but

---

*The point biserial correlations of the dichotomous criterion with the n individual variables and with the propinquity index are of course part of the correlation matrix and should therefore be based on the same sample.

also because of the bouncing beta problem and because the betas may bounce especially erratically when a variable involving squares is involved, such as the propinquity index, the sample used for the multiple correlation analysis should probably be substantially larger than either of the other two, to control as well as possible this instability.

3. The propinquity index should be retained in the final regression equation only for those dichotomous criteria where it seems to add significantly to the cross-validated multiple correlation. This is probably most likely to occur in situations where both overqualification and underqualification are serious deterrents to group membership - rather than the somewhat more usual situation where underqualification is an overwhelmingly more potent deterrent than overqualification is. In the vocational planning or vocational guidance situation this is more likely to happen in the case of middle-level occupations, and perhaps even, to a certain extent, in the case of lower-level occupations, than in the case of occupations that demand very high levels of ability. In the case of these more demanding occupations, the higher one's qualifications are the better.

CONCLUSIONS

In summary, the propinquity index may be used in two ways in helping the individual develop his educational and vocational plans. Converted to a propinquile, it can constitute one of many separate items of information used in arriving at important decisions. Or in many circumstances the propinquity index may function better as one of the predictors in a multiple regression equation to predict a dichotomous group membership criterion.

At the risk of oversimplifying horribly - and I know that I am - I would like to close by suggesting that perhaps the principal statistical approaches that should be used in three fields of application - selection, classification, and guidance - are as follows:

For personnel classification:	Multiple discriminant function analysis.
For personnel selection:	Multiple correlation approach.
For guidance:	Propinquiles, and the multiple point biserial R approach, modified to include the propinquity index as one of the predictors.

REFERENCES

Rulon, P. J. "The Stanine and the Separile: A Fable", Personnel Psychology, 4, 1951, 99-114.

Rulon, P. J. "Distinctions between Discriminant and Regression Analyses and a Geometric Interpretation of the Discriminant Function", in "The Multiple Discriminant Function - A Symposium". Harvard Educational Review, vol. XXI, No. 2, Spring 1951, 80-90.

Tiedeman, D. V. "The Utility of the Discriminant Function in Psychological and Guidance Investigations", in "The Multiple Discriminant Function - A Symposium", Harvard Educational Review, vol. XXI, No. 2, Spring 1951,71-80.

APPENDIX

I. NOTATION

n = no. of variables (= no. of dimensions)

g = no. of groups

N_j = no. of cases in group j

N = total no. of cases

$$N = \sum_{j=1}^{g} N_j \tag{1}$$

X_{ik} = raw score of individual k on variable i

$$i = 1,2,3,\ldots n$$
$$k = 1,2,3,\ldots N_j$$

$\bar{X}_1$ = mean of variable X_i for total group

$$\bar{X}_i = \frac{\sum_{k=1}^{N} X_{ik}}{N} \tag{2}$$

s_{x_i} = sample standard deviation of variable X_i for total group

$$s_{x_i} = \sqrt{\frac{\sum_{k=1}^{N} (X_{ik} - \overline{X}_i)^2}{N}} \quad (3)$$

σ_{x_i} = corresponding estimate of population standard deviation

$$\sigma_{x_i} = s_{x_i} \sqrt{\frac{N}{N-1}} \quad (4)$$

z_{ik} = standard score of individual k on variable i

$$z_{ik} = \frac{X_{ik} - \overline{X}_i}{\sigma_{x_i}} \quad (5)$$

$$\overline{z}_i = 0 \quad (6)$$

Note that σ_x not s_x is in the denominator of z_{ik} in this algebraic development.

$$\therefore \quad s_{z_i} = \sqrt{\frac{N-1}{N}} \quad (7)$$

$$\sigma_{z_i} = 1 \quad (8)$$

$\overline{X}_{ij}$ = means of variable X_i for group j

$$\overline{X}_{ij} = \frac{\sum_{k=1}^{N_j} X_{ik}}{N_j} \qquad (9)$$

$i = 1,2,3,\ldots n$
$j = 1,2,3,\ldots g$

$s_{x_{ij}}$ = sample standard deviation of variable i for group j

$$s_{x_{ij}} = \sqrt{\frac{\sum_{k=1}^{N_j} (X_{ik} - \overline{X}_{ij})^2}{N_j}} \qquad (10)$$

$\sigma_{x_{ij}}$ = estimate of population standard deviation of variable i for group j

$$\sigma_{x_{ij}} = s_{x_{ij}} \sqrt{\frac{N_j}{N_j - 1}} \qquad (11)$$

U_{ijk} = propinquity component indicating how much like group j individual k is, with respect to variable i.

w_{ij} = weight representing the relevance of component U_i as an indicator of membership of individual in group j.

δ_{jk} = propinquity index for individual k indicating how much like group j he is.

P_{jk} = propinquile for individual k corresponding to his propinquity index for group j. The propinquile is a percentile based on the individuals in group j. A propinquile of 100 corresponds to a propinquity index of 0.

II. FORMULAS

$$U_{ijk} = \left(\frac{z_{ik} - \bar{z}_{ij}}{\sigma_{z_{ij}}} \right)^2 \qquad (12)$$

$$= \left(\frac{X_{ik} - \bar{X}_{ij}}{\sigma_{x_{ij}}} \right)^2 \qquad (13)$$

$$\delta_{jk} = \sqrt{\sum_{i=1}^{n} w_{ij} \, U_{ijk}} \qquad (14)$$

$$\sigma_{z_{ij}} = \frac{\sigma_{x_{ij}}}{\sigma_{x_i}} \qquad (15)$$

$$= \frac{\sigma_{x_{ij}}}{s_{x_i}} \sqrt{\frac{N - 1}{N}} \qquad (16)$$

Formulas for intermediate values in computing w_{ij}

$$w'_1 = \log_{10} \frac{1}{\sigma_{z_{ij}}} \tag{17}$$

$$w'_2 = \frac{1}{\sigma_{z_{ij}}} - 1 \tag{18}$$

$$w'_3 = 1 - \sigma_{z_{ij}} \tag{19}$$

$$w'_4 = \frac{1}{\sigma^2_{z_{ij}}} - 1 \tag{20}$$

$$w'_5 = 1 - \sigma^2_{z_{ij}} \tag{21}$$

Formulas for w_{ij}

$w_1 = w'_1$ with negative values changed to 0 (22)

$w_2 = w'_2$ " " " " " " (23)

$w_3 = w'_3$ " " " " " " (24)

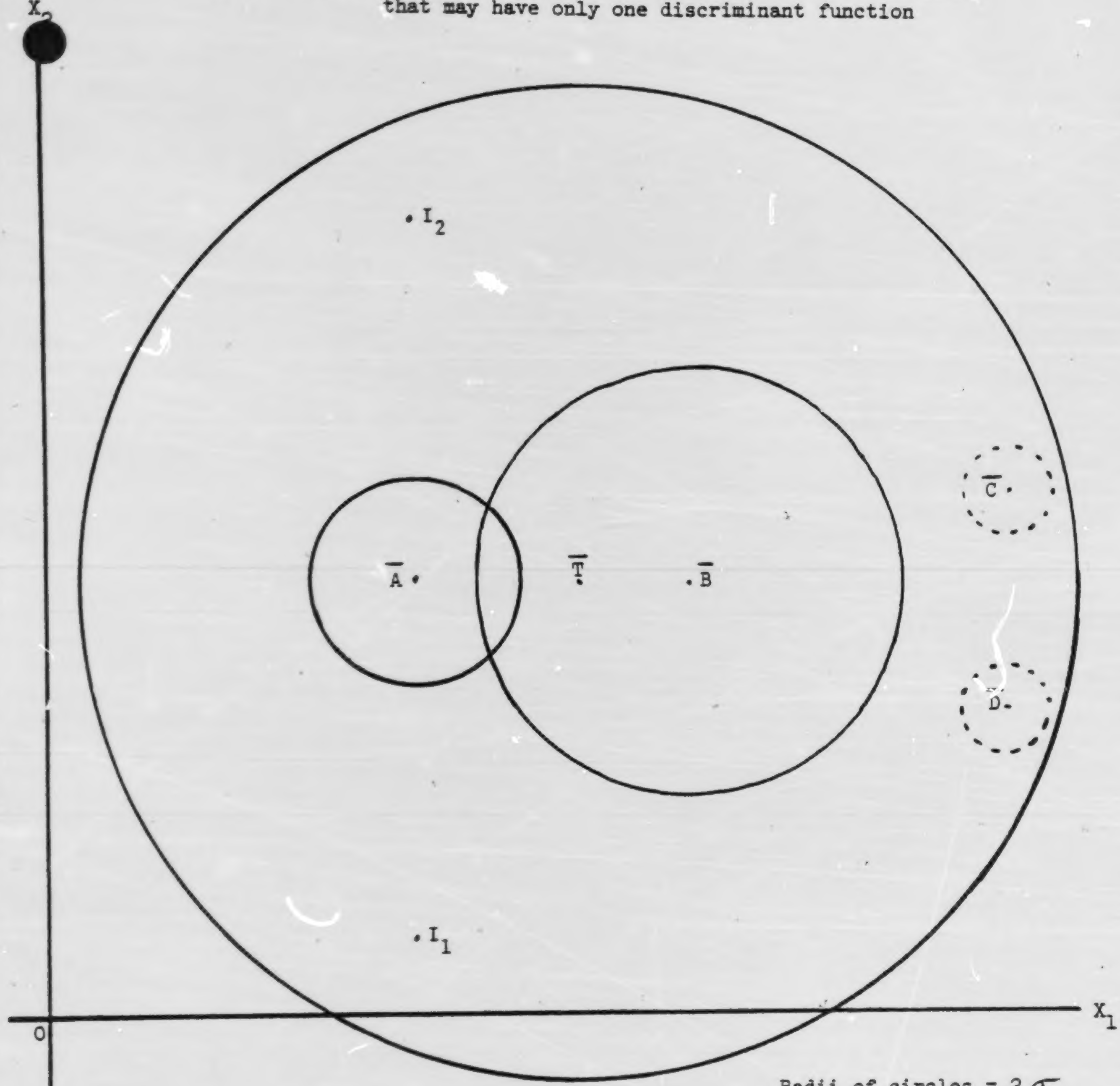
$w_4 = w'_4$ " " " " " " (25)

$w_5 = w'_5$ " " " " " " (26)

$$w_6 = w^2_2 \tag{27}$$

$$w_7 = w^2_3 \tag{28}$$

FIGURE 1. Illustrating data with two important dimensions that may have only one discriminant function



Radii of circles = 2 σ

Groups A, B, and T-A-B have only one discriminant function. When Groups C and D are added, a second discriminant function appears.

I_1 and I_2 are individuals classified on the basis of discriminant analysis as perfect fits in Group A (unless Groups C and D have been included in the analysis), but propinquity analysis definitely excludes them from Group A.

G1-20 3/10/70

PROPINQUITY ANALYSIS: A NEW GUIDANCE TOOL*

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

A multiple aptitude battery is a useful tool in vocational and educational guidance but there has been no clear agreement on the best way to use it. Assuming that the problem is how best to help the individual decide which of several possible occupational categories or educational programs is the one which will be best for him, three statistical techniques for developing a predictive equation for each category have been available: (1) multiple correlation other than multiple point biserial, (2) multiple discriminant analysis, and (3) multiple point biserial correlation. In the interests of brevity, the first-named of those three techniques will be referred to in this paper simply as "multiple correlation".

The term "multiple correlation technique", as used in this context, means that a separate within-group criterion of effectiveness of performance is available and that for prediction of membership in each category a separate multiple correlation coefficient and multiple regression equation are computed.

The second technique, multiple discriminant analysis, is applicable when the sole criterion available is membership in a group; in other words when there is no within-group criterion to distinguish among members of the group. The technique produces a set of discriminant functions, each of which is a linear composite of the original predictor variables. The discriminant functions have to be considered in combination, since they do not correspond on a one-to-one basis to the criterion categories.

---

*A very slightly abridged version of this paper was presented at the Annual Meeting of the National Council on Measurement in Education, in Minneapolis, March 6, 1970.

The third commonly used approach, multiple point biserial correlation, is really a cross between the first two. It is multiple correlation against a dichotomous criterion, and therefore, like the multiple discriminant analysis technique, it is usable even when membership-versus-nonmembership in the group is the only criterion information available. The resulting multiple regression function is exactly equivalent (except for a scaling factor) to a discriminant function that discriminates optimally between the group in question and all other groups combined.

But even the simultaneous use of all three of these techniques -- multiple correlation (against a continuous criterion), multiple discriminant analysis, and multiple point biserial R -- doesn't yield all the relevant information available. None of the three really comes to grips adequately with the kind of situation, sometimes occurring in the use of aptitude batteries, in which overqualification for group membership is just about as undesirable as underqualification is. The problem is still further aggravated when, as often happens, group membership criteria are the only kind available. Other difficulties, particularly applicable to the discriminant function approach, lie in the fact that because most of the discriminant functions are bipolar, they are hard to interpret, hard to explain, and generally obscure in meaning. This makes them peculiarly unsuitable for use in guidance. And discriminant functions have the further disadvantage that the results for one criterion category depend to an inordinate degree on what other categories happen to be included in the analysis.

In an effort to avoid all these problems and to supplement information provided by the more usual approaches, the writer has worked out a new approach, which is still in the developmental stage. This approach, called propinquity analysis, requires no criteria other than group membership, and although it superficially has some of the features of each of the other approaches it is actually quite different from any of them.*

THE NATURE AND CHARACTERISTICS OF PROPINQUITY INDEXES

Let's assume we have a scatterplot in which each individual's set of scores is represented by a point in n-dimensional space, where the standard score scale on each test constitutes a separate dimension. Let's now suppose that for purposes of developing propinquity indexes for a given occupational group the scatterplot is rescaled, weighting each dimension by a value representing, at least approximately, the relevance of the corresponding variable in identifying group members. An individual's propinquity index with respect to that group may then be defined as his geometric distance from the group centroid, in this rescaled space. A minus sign is attached to the distance, so that 0 is the maximum value of the index. A zero index indicates that the individual's scores on relevant variables coincide with the group centroid. The greater the distance, the larger the absolute value of the index. The purpose of the minus sign, therefore, is to orient the propinquity index properly, so that when it is used in a correlation matrix there won't be inconvenient negative correlations. The higher the algebraic value of the index, the closer the individual is to the centroid. (Hence the term "propinquity".)

Formula 14 (or 15)** represents the squared distance in regard to a single variable. Formula 16 gives the propinquity index, δ, with w as the weight representing the relevance of a particular variable.

Formulas 27-33 are seven different formulas giving different results, that might be used for determining the w's. All these formulas have the desired characteristic of giving a weight of 0 for irrelevant variables and a positive weight for relevant variables. There are no negative weights. The indicator of relevance for a variable is a function of the ratio of the standard deviation within the group to the standard deviation of the total sample. Thus there is an assumption that one way a variable's relevance may be manifested is by a somewhat restricted range, and that the more relevant the variable, the more restricted the range. This phenomenon may be seriously affected when some of the variables have skewed distributions. For such variables the ratio of group σ to total σ may be erratic, resulting in peculiar weights. Therefore it is desirable that all variables used in a propinquity analysis have normal distributions. If there is no reason to believe that the distributions are at least approximately normal they should be normalized.

---

*An earlier paper by Shaycoft entitled "A New Multivariate Index for Use in Educational Planning", which was presented at the American Psychological Association Convention in Washington on September 2, 1969, goes into considerably more detail on the limitations of the other three methods. Copies are available on request, while the supply lasts.

**All formulas in this paper are in the Appendix, which also contains a section defining all the notation used.

Multiple regression weights (against the dichotomous criterion of group membership), as suggested by formula 39, may constitute a more effective approach than any of the seven formulas 27-33. Regression weights would of course have to be cross-validated; and small negative weights should probably be changed to 0 first. (Presumably there wouldn't be any large negative weights.) The chief disadvantage of regression weights lies in the very practical consideration that even with a high-powered computer, the computation of the betas for these squared terms is an extraordinarily complex, time-consuming, and expensive operation when the number of cases, number of variables, and number of groups are all comparatively large (as is true in the case of the Project TALENT data, to which this technique is being applied).

Because the variables used in computing the propinquity index are raw scores rather than principal components, discriminant functions, or some other kind of uncorrelated variables, propinquity indexes may lack some of the mathematical precision and invariance of statistics computed in a geometric space where orthogonal axes correspond to uncorrelated variables. But this seems a small price to pay for the twin advantages of interpretability and ready explainability.

PROPINQUILES

Now we come to a problem in connection with propinquity indexes. When the number of variables entering into the index is greater than one, the distribution is not normal and its basic shape varies with the number of variables involved. Propinquity indexes based on three variables have an entirely different distribution from those based on two, and so forth.

And even when propinquity indexes for a set of groups are all based on the same battery they still may not be directly comparable, because though all indexes technically are based on the same number of variables, some variables may have zero weights for certain groups so that the number of dimensions is in effect different for different groups. Since propinquity indexes for different categories therefore are not generally directly comparable, they have to be converted to some uniform scale in order to be compared - and percentiles serve this purpose effectively and directly. This calls for coinage of a new word. The percentiles are called "propinquiles".

PROBLEMS IN DETERMINING WEIGHTS

The nature of the change in the shape of the distribution of the propinquity index as the number of variables increases causes some of the difficulties in determining a wholly suitable a priori formula for the weights. In a small-scale tryout of formulas 27-33, for eight career groups, results suggested formula 27 was slightly better than any of the others; for six of the eight groups, it produced δ^2 values that had at least slightly higher point biserial correlations with the group membership criterion than did any of the other formulas. In addition to this rather slight empirical evidence there are some theoretical arguments in support of formula 27 — for instance the fact that it bears a rather close relationship to the square of a type of correlation ratio. Nevertheless even this "best" a priori formula gives somewhat disappointing results, leading to the suspicion that it may not be the best weighting procedure after all.

Possibly multiple regression weights should be used, rather than any *a priori* formula, *despite* the computational complexities that this would introduce into the data analysis.

A possible alternative is the so-called city block model for defining distance. This alternative kind of propinquity index is represented by δ', and its value is given by formula 21. With it go new formulas w_1'-w_5' for *a priori* weights (Formulas 34-38). This alternative system, δ',hasn't been tried yet, and it is probably worth trying, but the chances seem slight that it will prove to have any advantages over δ, while δ has at least one major advantage over δ' -- the fact that it is invariant under rotation of axes.

This brings us to the empirical data summarized in Table 1 (in the Appendix). It is important to bear in mind that this is *preliminary* data. The study is still in progress.

EMPIRICAL DATA

Table 1 is based on about 14000 twelfth-grade boys tested in Project TALENT in 1960 for whom scores on all 109 predictor variables used in the study are available and for whom follow-up data obtained five years after the class graduated from high school are also available. These follow-up data provided the information about the long-range career plans of the TALENT sample. Ten of these career-plan groups have been selected for inclusion in Table 1.*

The total group has been divided into two parallel samples, A and B, so that propinquity indexes could be calculated for the individuals in one sample on the basis of statistics (means and standard deviations) determined from the other, in order to avoid capitalizing on chance.

All 109 variables were normalized with a mean of 0 and standard deviation of 1, on the basis of Sample A distributions. The same normalization conversion table was then applied to the Sample B cases, to make the Sample B distributions approximately normal. Propinquity indexes were calculated for every Sample B case, using weights, means, and standard deviations calculated from Sample A data. Formula 27 was used for the weights.

---

*The choice of these particular 10 career plan groups was based primarily on their size (they are all large enough to provide comparatively stable data) and the variety of fields and levels they represent.

Columns 1-3 show the number of cases in each group. Column 4 shows the proportion, and column 5 shows the factor by which a point biserial correlation is multiplied to convert it to a biserial. Column 6 contains the conventional multiple point biserial R, with group membership as the dichotomous criterion and the 109 normalized variables as the predictors. These point biserials are necessarily quite low because of the very low ceiling on point biserial correlations when the dichotomous split is far out on a tail of the hypothesized normal distribution. The conversion to multiple biserial correlations, shown in column 7, makes all ten values directly comparable, since it eliminates the effect of different numbers of cases in the different groups.

Interpretation of the biserial correlations of course requires one to conceptualize that underlying each dichotomy is a unidimensional normal continuum; for instance that there is a normally distributed continuum of "CPA-ness" — tendency to become a CPA — and a threshhold above which everybody becomes a CPA and below which nobody does. It is important to bear in mind that this hypothesized threshhold applies only to the hypothesized normal variable underlying the dichotomized criterion -- and not to any composite of the predictive variables (except in the extremely improbable event that the composite has perfect validity).

Column 8 contains the point biserial correlation of the propinquity index, δ, with the group membership criterion, and column 9 contains the corresponding biserial. These correlations are substantial; they are lower than the corresponding multiples in columns 6 and 7 but that is to be expected since the correlations in columns 6 and 7 are based on 109 variables combined optimally and are not cross-validation values -- so there may be a little capitalization on chance there, while there is no capitalization on chance in the case of the propinquity indexes.

The critical question is whether adding the propinquity index as a 110th predictor would increase the column 6 multiple point biserial correlations significantly. The data for a definite answer on that question are not yet available, but preliminary indications are that though the amount of increase produced in the R by the addition of the propinquity index to the group of predictive variables would be relatively small (assuming that formula 27 is used for the weights) it would be significant. To the extent that that δ adds new predictive variance, not measured by the 109 predictive variables used for the column 6 multiples, R will be raised by the addition of δ as a predictor. And there is evidence that much of the propinquity index variance is unique. For instance the variables that are weighted the highest in computing δ are not necessarily the ones on which group members score especially high or especially low. They may be variables on which the average score of a group member is very close to the overall mean and the point biserial of that variable with the group membership criterion is probably correspondingly low.

The last eight columns of the table provide some evidence on this point. The highest and second highest weights for the computation of each of the 10 propinquity indexes are shown in columns 13 and 17. The standard score means and point biserials for many of these variables are quite close to 0, as can be seen in columns 15, 16, 19, and 20.

And as has already been indicated, the empirical data provide some suggestion that using multiple regression weights or some other modification of the formulas might make a substantial improvement in the results. One indication of this is in columns 10-12. Some of the dichotomous criteria turn out to have higher correlations with the δ computed for another group than with the appropriate δ. This certainly strongly suggests that better weights could be found.

CONCLUSIONS

Propinquity analysis is not in any sense of the word a replacement for the multiple correlation approach, nor for the multiple point biserial R approach. Rather, it should serve as an adjunct to both these procedures. It may be used in two ways in helping the individual develop his educational and vocational plans. Converted to a propinquile, it can constitute one of many separate items of information used in arriving at important decisions. Or in many circumstances the propinquity index may function better as one of the predictors in a multiple regression equation to predict a dichotomous group membership criterion.*

In the vocational planning or vocational guidance situation propinquity analysis is likely to be most useful for middle-level occupations, and perhaps even, to a certain extent, for lower-level occupations, rather than for the more demanding occupations, where underqualification is an overwhelmingly more potent deterrent than overqualification is, and the higher one's qualifications are the better.

In any event, further exploration of computing procedures is necessary. It seems clear both on logical grounds and on the basis of preliminary data that the components entering into the computation of a propinquity index provide unique predictive variance — variance not provided by the conventional scores -- and that this is not just error variance; it is specific reliable variance, and when appropriately weighted it should contribute significantly to the validity of prediction. But better weighting procedures (and perhaps other computational modifications as well) must be uncovered, in order to mine efficiently this rich and hitherto unexploited source of valid variance.

---

*This of course couldn't be proposed if the propinquity index were merely some sort of linear function of the $\underline{n}$ initial predictors, but it isn't, since squared terms are also involved. The propinquity index thus is not linearly dependent on the $\underline{n}$ variables which enter into it.

SELECTED REFERENCES ON PROJECT TALENT

Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Gorham, W. A., Orr, D. B., & Goldberg, I. Design for a Study of American Youth, 248 pp. Boston: Houghton Mifflin, 1962.

Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., & Goldberg, I. Studies of the American High School. (Cooperative Research Project No. 226) Project TALENT Office, 1962. 375 pp.

Flanagan, J.C., Davis, F.B., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., Goldberg, I., & Neyman, C.A., Jr. The American High School Student. (Cooperative Research Project No. 635) Project TALENT Office, 1964. 738 pp.

Flanagan, J.C., Cooley, W.W., Lohnes, P.R., Schoenfeldt, L.F., Holdeman, R.W., Combs, Janet, & Becker, Susan. Project TALENT One-Year Follow-Up Studies. (Cooperative Research Project No. 2333) Project TALENT Office, 1966. 348 pp.

Shaycoft, Marion F., Dailey, J.T., Orr, D.B., Neyman, C.A., Jr., & Sherman, S.E. Studies of a Complete Age Group--Age 15. (Cooperative Research Project No. 566) Project TALENT Office, 1963. 288 pp.

Shaycoft, Marion F. The High School Years: Growth in Cognitive Skills. (Cooperative Research Project No. 3051) Project TALENT Office, 1967. 376 pp.

APPENDIX

I. NOTATION

n = no. of variables (= no. of dimensions)

g = no. of groups

N_j = no. of cases in group j

N = total no. of cases

$$N = \sum_{j=1}^{g} N_j \tag{1}$$

p_j = proportion of cases in group j

$$p_j = {}^{N_j}/_{N} \tag{2}$$

$$q_j = 1 - p_j \tag{3}$$

y_j = normal ordinate corresponding to point of split between p_j and q_j

X_{ik} = raw score (or normalized score if the variables have been normalized) of individual k on variable i

$$i = 1,2,3,\ldots n$$
$$k = 1,2,3,\ldots N_j$$

$\bar{X}_1$ = mean of variable X_i for total group

$$\bar{X}_i = \frac{\sum_{k=1}^{N} X_{ik}}{N} \tag{4}$$

s_{x_i} = sample standard deviation of variable X_i for total group

$$s_{x_i} = \sqrt{\frac{\sum_{k=1}^{N} (X_{ik} - \overline{X}_i)^2}{N}} \qquad (5)$$

σ_{x_i} = corresponding estimate of population standard deviation

$$\sigma_{x_i} = s_{x_i} \sqrt{\frac{N}{N-1}} \qquad (6)$$

z_{ik} = standard score of individual k on variable i

$$z_{ik} = \frac{X_{ik} - \overline{X}_i}{\sigma_{x_i}} \qquad (7)$$

$$\overline{z}_i = 0 \qquad (8)$$

Note that σ_x not s_x is in the denominator of z_{ik} in this algebraic development.

$$\therefore \quad s_{z_i} = \sqrt{\frac{N-1}{N}} \qquad (9)$$

$$\sigma_{z_i} = 1 \qquad (10)$$

$\overline{X}_{ij}$ = means of variable X_i for group j

$$\overline{X}_{ij} = \frac{\sum_{k=1}^{N_j} X_{ik}}{N_j} \qquad (11)$$

i = 1,2,3,...n
j = 1,2,3,...g

$s_{x_{ij}}$ = sample standard deviation of variable i for group j

$$s_{x_{ij}} = \sqrt{\frac{\sum_{k=1}^{N_j} (X_{ik} - \overline{X}_{ij})^2}{N_j}} \qquad (12)$$

$\sigma_{x_{ij}}$ = estimate of population standard deviation of variable i for group j

$$\sigma_{x_{ij}} = s_{x_{ij}} \sqrt{\frac{N_j}{N_j - 1}} \qquad (13)$$

U_{ijk} = propinquity component indicating how much like group j individual k is, with respect to variable i.

L_{ijk} is an alternative to U_{ijk}

w_{ij} = weight representing the relevance of component U_i as an indicator of membership of individual in group j.

w'_{ij} = weight representing the relevance of component L_i as an indicator of membership of individual in group j.

δ_{jk} = propinquity index for individual k indicating how much like group j he is, assuming the U_{ijk} formula has been used for defining likeness.

δ'_{jk} = propinquity index for individual k indicating how much like group j he is, assuming the L_{ijk} formula has been used for defining likeness.

P_{jk} = propinquile for individual k corresponding to his propinquity index for group j. The propinquile is a percentile based on the individuals in group j. A propinquile of 100 corresponds to a propinquity index of 0.

II. FORMULAS

$$U_{ijk} = \left(\frac{z_{ik} - \bar{z}_{ij}}{\sigma_{z_{ij}}} \right)^2 \tag{14}$$

$$= \left(\frac{X_{ik} - \bar{X}_{ij}}{\sigma_{x_{ij}}} \right)^2 \tag{15}$$

$$\delta_{jk} = -\sqrt{\sum_{i=1}^{n} w_{ij} U_{ijk}} \tag{16}$$

$$\sigma_{z_{ij}} = \frac{\sigma_{x_{ij}}}{\sigma_{x_i}} \tag{17}$$

$$= \frac{\sigma_{x_{ij}}}{s_{x_i}} \sqrt{\frac{N - 1}{N}} \tag{18}$$

$$L_{ijk} = \sqrt{U_{ijk}} \tag{19}$$

$$= \left| \frac{X_{ik} - \bar{X}_{ij}}{\sigma_{x_{ij}}} \right| \tag{20}$$

$$\delta'_{jk} = -\sum_{i=1}^{n} w'_{ij} L_{ijk} \tag{21}$$

Formulas for intermediate values in computing w_{ij} and w'_{ij}

$$w_1'' = 1 - \sigma^2_{z_{ij}} \qquad (22)$$

$$w_2'' = \frac{1}{\sigma^2_{z_{ij}}} - 1 \qquad (23)$$

$$w_3'' = 1 - \sigma_{z_{ij}} \qquad (24)$$

$$w_4'' = \frac{1}{\sigma_{z_{ij}}} - 1 \qquad (25)$$

$$w_5'' = \log_{10} \frac{1}{\sigma_{z_{ij}}} \qquad (26)$$

Formulas for w_{ij} and w'_{ij}

$w_1 = w_1''$ with negative values changed to 0 (27)
$w_2 = w_2''$ " " " " " " (28)
$w_3 = w_3''$ " " " " " " (29)
$w_4 = w_4''$ " " " " " " (30)
$w_5 = w_5''$ " " " " " " (31)

$$w_6 = {w_3}^2 \qquad (32)$$

$$w_7 = {w_4}^2 \qquad (33)$$

$w_1' = \sqrt{w_1}$ (34)
$w_2' = \sqrt{w_2}$ (35)
$w_3' = w_3$ (36)
$w_4' = w_4$ (37)
$w_5' = w_5$ (38)

w_o = multiple regression beta weight when the predictors are U-variables (39)

w_o' = multiple regression beta weight when the predictors are L-variables (40)

TABLE 1: Multiple point biserial R's with group membership, point biserial r's with δ, corresponding biserial correlations, and related data, for 10 selected "purified" career groups

Based on five-year follow-up of twelfth-grade boys*; (N = 14123)

n = 109
Predictor variables: 109 normalized** variables from the TALENT Battery

j	Career group code #	Career	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			No. of cases							r with corresp δ		For highest $\lvert r_{j\delta} \rvert$ among the 10 for group j			For weights (w_{ij}) for computing appropriate δ***							
			A	B	T	p	$\frac{\sqrt{pq}}{y}$	Mult. pt. bis. R	Mult. bis. R	Pt. bis. $r_{j\delta}$	Bis $r_{j\delta}$	Pt. bis. $r_{j\delta}$	Bis. $r_{j\delta}$	Group Code for δ	Highest w_{ij}	Corresp. variable (i)	$\bar{z}_{ij}$	Bis. r_{ji}	Second highest w_{ij}	Corresp. variable (i)	$\bar{z}_{ij}$	Bis. r_{ji}
1	748	Salesman	175	175	350	.02494	2.6734	.156	.417	.060	.162	(.060)	(.162)	(748)	.3482	R-142	-.070	-.040	.3112	P-820	-.330	-.059
2	730	Business mgmt.	256	255	511	.03634	2.3482	.167	.391	.057	.133	.072	.169	393	.2994	A-410	.190	.085	.2791	P-702	.134	.030
3	716	CPA	140	140	280	.01995	2.8949	.183	.530	.067	.193	(.067)	(.193)	(716)	.5124	R-101	.122	.046	.4168	R-311	.306	.055
4	393	Lawyer	197	196	393	.02793	2.5697	.259	.665	.112	.289	-.149	-.383	639	.5633	R-136	.473	.267	.4782	R-101	.110	.069
5	329	M.D. (NEC)	115	114	229	.01625	3.1174	.281	.876	.120	.374	-.179	-.557	899	.6794	R-136	.590	.212	.6049	R-250	1.002	.365
6	242	Elec. engineer	122	121	243	.01724	3.0504	.246	.749	.108	.331	-.123	-.376	899	.4973	R-270	.756	.336	.4779	P-701	1.177	.448
7	811	Electronic tech.	106	106	212	.01511	3.2010	.178	.570	.049	.156	(.049)	(.156)	(811)	.4287	R-234	.028	-.115	.3984	P-714	.519	.205
8	125	Electronics (NEC)	78	78	156	.01112	3.5874	.166	.594	.033	.120	.034	.123	811	.4122	R-250	-.025	-.097	.3991	R-232	-.294	-.072
9	639	Farming (NEC)	122	122	244	.01739	3.0414	.278	.846	.104	.315	-.131	-.398	329	.4639	R-104	-.748	-.286	.4573	R-250	-.630	-.316
10	899	Unskilled labor	139	138	277	.01967	2.9087	.195	.567	.112	.325	-.141	-.410	329	.5486	R-333	-.770	-.279	.4857	R-106	-.888	-.335
⋮																						
Total			7106	7017	14123	1.00000																
Foot note #'s	1	2	3	4		4	4	4,6-7		4-6				1	3,8	9	3	4,6	3,8	9	3	4,6

* The group was divided into 2 samples, A and B; sample A consisting of the 1st, 3rd, 5th,... cases in each career group, and sample B consisting of the 2nd, 4th, 6th....

** Normalization was based on the sample A distribution.

*** δ is the propinquity index to which column 8 (or 9) applies.

1. TALENT career plan code series Q-2-cp-A

2. NEC stands for "Not elsewhere classified".

3. Based on sample A

4. Based on sample B

5. $\bar{X}$ and σ values entering into the computation of w_{ij} (Formula 27) and δ_j are based on sample A; 109 U-variables based on the normalized TALENT variables are used.

6. Dichotomous group membership variable is the criterion (X_j).

7. The 109 normalized TALENT variables are the predictors.

8. Excluding w_{ij}'s for dichotomous predictor variables.

9. Code for TALENT variables represented in columns 14 and 18

R-101	Screening
R-104	Music Information
R-106	Math Information
R-136	Journalism Information
R-142	Bible Information
R-232	Capitalization (English subtest)
R-234	English Usage (English subtest)
R-250	Reading Comprehension
R-270	Mechanical Reasoning
R-311	Arith. Reasoning
R-333	Adv. h.s. math.
A-410	Arith Comp. attempts
P-701	Interest in phys. sci., engineering, math.
P-702	Interest in biol. sci. and medicine
P-714	Mechanical-Technical interest
P-820	High school grades

Appendix G2

THE PROJECT TALENT APPROACH TO THE CAREER-GROUPING PROBLEM: CUTTING THE GORDIAN KNOT*

Marion F. Shaycoft
American Institutes for Research

For many years researchers concerned with career guidance and career planning have been trying to develop or use statistical methodologies applicable to the problem of grouping careers. The goal has been to establish career groupings, and principles by which each kind of job or career can be assigned to one and only one of the groups thus established. The purpose--and it is a very important one--has been to facilitate career guidance by cutting the number of alternatives from several thousand (the number of different kinds of jobs) down to a much smaller and more manageable number (the number of career groups, each group consisting of many different kinds of jobs).

Ideally, of course, from the viewpoint of their use in vocational guidance all the kinds of jobs or careers in a career group should be very similar in the sense that they require the same pattern of aptitudes and traits of temperament, and much the same preparation. Such classification of jobs into fairly homogeneous categories for purposes of career guidance is useful because it maximizes flexibility. Boys and girls of high school age aren't channeled into a narrow pathway that leads to one and only one kind of job. Instead, they maintain as many options as possible as long as possible.

Researchers into the career categorization problem have used a variety of multivariate analysis and other techniques in their explorations--various kinds of profile analysis, multiple discriminant analysis, and hierarchical analysis, to name just a few.

We on Project TALENT are among those who have used hierarchical analysis** to explore the career-categorization problem. Hierarchical analysis is a multivariate analysis procedure designed to reveal the "natural groups" into which the entities being classified fall--if such natural groups in fact exist.

---

*Paper presented at the Eighth Invitational Conference on Systems Under Construction in Career Education and Development, A.I.R., Palo Alto, October 1971.

**Flanagan, J.C., Shaycoft, Marion F., Richards, J.M., Jr., and Claudy, J.G. Five Years after High School. Palo Alto, Calif.: American Institutes for Research, 1971. Chapter 5.

A word or two about Project TALENT is in order at this point. In 1960 several hundred thousand high school students--all the 9th-, 10th-, 11th-, and 12th-graders in a stratified random sample of all the secondary schools (parochial and private as well as public) in the United States--were given a comprehensive two-day battery of tests measuring aptitudes, achievement levels, and patterns of information, and questionnaires (inventories) to secure information about their interests, traits of temperament, and background characteristics. The purposes of the project, as originally stated, have been:

1. To obtain a national inventory of human resources.
2. To develop a set of standards for educational and psychological measurements.
3. To provide a comprehensive counseling guide indicating the patterns of aptitude and ability which are predictive of success and satisfaction in various careers.
4. To formulate a better understanding of how young people choose their life work.
5. To identify the educational experiences which better prepare students for their life work.

In line with some of these purposes, particularly the last three, we have been following up the members of the sample on a long-term basis, to find out about their post-high-school educational activities, jobs, long-range career plans, and other matters related to their early adult years. Follow-up questionnaires were sent to members of the TALENT sample one year after high school, and again five years after; now we are following these same young men and young women up a third time, eleven years after high school.

In processing the returned questionnaires, the respondents' long-range career plans are classified into one of nearly 1000 categories, using a four-digit code for the purpose. This is a "hand operation," done by a crew of coders. Essentially the same set of codes and the same procedure are used for coding the respondents' current jobs. This 1000-group categorization system was designed specifically for Project TALENT. It is an expandable code, and obviously one that incorporates considerable detail. It is not identical with the coding system used by the Department of Labor in its Dictionary of Occupational Titles, nor with any other system previously in use. The Dictionary of Occupational Titles code serves admirably the purpose for which it was developed, but for our purposes in Project TALENT research different emphasis was needed, with far less detail in certain areas and somewhat more in others; hence the need to develop our own coding system. One of the special requirements of the Project TALENT code was that it had to provide the capability of handling incompletely developed career plans; e.g., something as vague as the statement by the respondent that he wants to "work with people"; or perhaps that he wants to "work for the government".

Since the nearly 1000 categories of the four-digit code are far more than can be handled conveniently in most data analyses it was recognized at the outset that some subsequent collapsing of categories would be necessary. Why, then, did we start out with a more detailed code than we were going to want to use in specific research investigations? The answer, of course, is that we wanted to allow maximum flexibility in combining categories later on, to permit different kinds and degrees of condensation of categories for different purposes. And as things turned out, that is exactly what happened; different researchers combined categories in different ways. The purpose of the present report is to discuss one line of research on the subject of combining categories. The first step, after all of the questionnaire responses, including the job and career-plan four-digit codes, had been put on computer tape, was to add a simplified categorization of jobs and career plans to the tape. This was accomplished by converting the thousand-category four-digit codes to a somewhat coarser grouping, containing only a couple of hundred categories and utilizing a three-digit coding system. These coarser categories were established on a judgmental basis, by combining some of the four-digit categories into a single category in the three-digit system. This was in fact not a very drastic condensation. It was regarded as merely the first step in reducing the number of categories for potential use in educational and career guidance of high school students. As has already been suggested, the intention was to undertake a program of exploratory research into ways of combining the groups further, to get a very much smaller number of groups.

One additional step was necessary, however, before this exploration could be undertaken. This step was "purification" of the career-plan groups. In using the TALENT Battery to predict success in a particular career field, a criterion is needed; but for many career fields five years after high school was a little too soon for prospective entrants even to have finished their training and actually started work in the field. And it was far too soon for them to have been in the field long enough for any criterion of quality of performance to be available. At this stage, then, the only practical criterion was the "group membership" criterion—i.e., the dichotomous variable indicating whether the person was or wasn't actually in the career field in question. Group membership is far from an ideal criterion; and it is certainly not an "ultimate criterion." But in view of the practical realities it was and is a useful proximal criterion, and has been utilized as such. And as a further acknowledgment of practical realities such as the lengthy training and preparation periods that are required for many careers, the definition of group membership was expanded to include not only those who had already obtained a job in the career field in question but also those who had realistic plans or expectations of doing so eventually. It was in regard to the matter of whether a person's career plan was reasonably realistic that "purification of career-plan groups" entered the picture. Individuals were eliminated from the "purified" groups if their follow-up data made it extremely unlikely that their career plans would ever be realized—for instance if a young man said he wanted to be a doctor but five years after high school hadn't even entered college yet and apparently had no plans to do so, he would be eliminated.

Hierarchical Analysis

The initial concept we started out exploring was that each occupational category (e.g., lawyer, doctor, welder, secretary, accountant, computer programer, psychologist, policeman, jockey, baseball player) falls naturally into one of a small number of natural clusters, these clusters being such that all occupations within a cluster are characterized by very similar patterns of aptitudes, abilities, interests, traits of temperament, and background, while between clusters there are substantial differences in these respects. Hierarchical analysis was the methodology chosen for use in the exploratory phase. On the basis of theoretical considerations it appeared that if any analytic approach were going to prove effective in establishing job families, hierarchical analysis would be that approach. In a hierarchical grouping procedure, one starts with a set of separate entities and the task is to decide how to combine them to get groups within which there will be considerable homogeneity but which will differ markedly from each other. In the case with which we are concerned the entities are job categories, or to be somewhat more precise, long-range career-plan categories since five years after high school not everyone had as yet actively entered the field which he planned eventually to make his career. These entities, career-plan categories, are described in terms of 109 numbers--the mean scores of the young men or young women in the category, on 109 variables (64 cognitive and 45 noncognitive) of the Project TALENT test battery. The hierarchical analysis procedure operates in a stepwise fashion, to combine entities whose index of difference is as low as possible or whose index of similarity is as high as possible. Difference may be expressed in a number of ways: as distance between entities, as dispersion of the combined group, or perhaps as a ratio of distance to dispersion. Likewise similarity may be expressed in numerous ways--as some measure of the amount of overlap between groups, for example. Thus the term hierarchical analysis doesn't cover just a single specific procedure. Rather it represents a whole family of procedures--so that the decision to use hierarchical analysis is just the starting point for a long series of methodological decisions. After careful consideration of the pros and cons of various possible procedures the hierarchical analyses were carried out in the following way:

1. Separate analyses were carried out for the 64 cognitive and the 45 noncognitive variables.
2. Orthogonal variables were used. Since the original 64 cognitive (or 45 noncognitive) variables were not all uncorrelated, and since it seemed desirable to use orthogonal variables in the hierarchical analysis, it was decided to do a complete principal components analysis of the 64 cognitive variables and a separate principal components analysis of the 45 noncognitive variables, and then to obtain each individual's scores on the principal components.*

---

*For this purpose each principal component was scaled to have a standard deviation of 1.

Principal component scores are suitable variables for use in the hierarchical analysis, since they are automatically orthogonal. Each individual was then hypothesized to be represented by a point plotted in 64-dimensional (or 45-dimensional) hyperspace, where the dimensions are the principal components. The location of the centroid* of each career-plan group was then determined.

3. Groups were then combined on the basis of how close together their centroids were. In addition to trying geometric distance between centroids as the basis for combining groups, five other functions, mostly variants of simple distance, were also tried.

The results were somewhat disappointing (at least they were disappointing initially) in the sense that they didn't conform to prior expectations. Hierarchical analysis just didn't reveal tight clusters of occupations, into which the groups fell naturally. Of the various formulas that were tried out, the simplest one, the simple distance formula, turned out to work about as well as any of the more complicated formulas and better than most. But none of them worked really well. The occupational groups just didn't combine neatly into tight clusters, and no formula, however complex and sophisticated, seemed about to make it appear that they did. With some modification of the procedure for interpreting the results--in other words modification of the set of rules for deciding where one group left off and a new one started--it was possible to develop some reasonable groups from the data; but to do this required the application of substantial doses of subjective judgment (which is really an inherent feature of the hierarchical analysis approach anyhow).

Table 1 shows the groups that were formed as a result of this modified hierarchical analysis and Table 2 shows the contents of each group--in other words the occupational categories included in each. These groupings have some utility, but they do not necessarily constitute the only possible useful way of doing the grouping. They represent, rather, just one of a potentially large number of equally satisfactory ways.

---

*The centroid is the point corresponding to the group's means on all 64 (or 45) principal components. (It should be noted that the use of centroids in this way automatically weights each principal component variable in accordance with its effectiveness in discriminating among groups. In other words dispersion among group means is far greater for those components that discriminate effectively among groups than for the ones that don't.)

TABLE 1. Job Families for Males*: Partial list
(Based on 5-year follow-up of grade 12 boys)

		No. of career groups included	No. of cases included
A.	Airplane pilot	1	99
B.	Business and industry	19	2885
C.	Architect	1	58
D.	Engineering and applied physical sciences	12	1108
E.	Math and physical science: quantitatively oriented professions	5	287
F.	Biological sciences: theoretical and applied	6	154
G.	"People-oriented" professions in the sciences	4	500
H.	Professions in the social sciences	4	763
I.	College professor: English	1	57
J.	Clergyman	1	176
K.	Teaching and other "people-oriented" professions (non-science)	6	734
L.	High school math teacher	1	62
M.	High school science teacher	1	50
N.	High school physical education teacher	1	57
O.	Miscellaneous skilled occupations	7	673
P.	Technician	6	610
Q.	Miscellaneous "blue-collar" jobs	18	1543
R.	Farming	2	372
S.	Protective	4	259
Y.	(Undecided)	-	803
		100**	11250***

---

*Based on two hierarchical analyses, one in 64 dimensions (principal components of 64 cognitive variables) and the other in 45 dimensions (principal components of 45 noncognitive variables).

This table includes all groups having at least 50 cases (in the five-year follow-up of grade 12 boys). Some smaller groups are also included which fit into a family defined by the larger groups.

**Includes 7 duplications; therefore 93 separate categories.

***Includes 756 duplications; therefore 10494 separate cases.

TABLE 2. Present composition of the job families

Job Family	Career Code	Career***	No. of cases
A.	841	Airplane pilot	99
B.	112	In business for self (NEC)	278
	120	Industry or business (NEC)	141
	711(2)	Banking and finance	116
	716	CPA	280
	717	Accountant, auditor, comptroller (exc. CPA)	202
	723	Efficiency expert, industrial engineer, production management	111
	730	Business management, business administration (NEC)	511
	731	Manufacturing management	95
	732	Wholesale or retail trade management; marketing	195
	746	Insurance salesman	61
	748	Salesman (NEC)	350
	749	Sales manager	73
	726	Personnel administration	81
	335	Pharmacist	91
	222*	Computer programer	121
	642*	U.S. Armed Forces: officer	108
	747**	Auto salesman	22
	743**	Stockbroker	23
	745**	Real estate sales	26
C.	250	Architect	58
D.	240	Engineer (NEC)	167
	241	Civil and/or hydraulic engineer	105
	242	Electrical and/or electronic engineer	243
	243	Mechanical or automotive engineer	143
	231*	Chemist	81
	642*	U.S. Armed Forces: officer	108
	222*	Computer programer	121
	244**	Aeronautical engineer	37
	245**	Chemical engineer	58
	234**	Geologist	17
	235**	Meteorologist	10
	230**	Scientist (NEC)	18
E.	211	Mathematician	46
	232	Physicist	64
	231*	Chemist	81
	462*	College professor: science	68
	461**	College professor: math	28

TABLE 2. (Continued)

Job Family	Career Code	Career***	No. of cases
F.	310	Biologist, zoologist, botanist, etc.	32
	314	Specialist in fish, wildlife, forestry, conservation, etc.	49
	313**	Specialist in agricultural science	35
	316**	Microbiologist	15
	317**	Biochemist	9
	322**	Surgeon	14
G.	329	Physician (NEC)	229
	332	Dentist	97
	360*	Psychologist	106
	462*	College professor: science	68
H.	360*	Psychologist	106
	393	Lawyer	393
	463	College professor: social studies	85
	460	College professor (NEC)	179
I.	464	College professor: English	57
J.	521	Clergyman	176
K.	400	Teaching (NEC)	423
	420	High School Teacher (NEC)	69
	423	High School Teacher: social studies	94
	450	School administration (principal, etc.): except college	43
	370	Social work	60
	412	Teaching elementary school	45
L.	421	High school math teacher	62
M.	422	High school science teacher	50
N.	429	High school physical education teacher	57
O.	861	Printing trades	106
	758	Computer operator, etc.	78
	798	Miscellaneous administrative	38
	797	Miscellaneous clerical	112
	864	Draftsman	101
	738	Supervisor (in a business)	87
	661*	Police (public)	151

TABLE 2. (Continued)

Job Family	Career Code	Career***	No. of cases
P.	811	Electronic technician	212
	125	Electronics (NEC)	156
	299	Lab technician, research asst., etc. (in physical sciences)	78
	222*	Computer programer	121
	347**	Medical and dental technician; technician in biol. and clinical sciences	16
	348**	Medical and dental technologist	27
Q.	102	Foreman (NEC)	49
	810	Electrician (NEC)	168
	812	Appliance repair	55
	820	Mechanic (NEC)	116
	821	Auto mechanic	74
	822	Airplane mechanic	68
	828	Machinist	176
	832	Carpenter	55
	833	Metal trades	92
	834	Bricklayer, mason roofer, printer, plasterer, etc.	62
	835	Plumber, pipefitter	72
	837	Misc. building and construction	71
	899	General labor (unspecialized)	277
	853	Auto, bus, and truck driver	85
	824**	Industrial machine repair	21
	813**	Phone installation, repair, maintenance	39
	838**	Mining, quarrying, well-drilling	16
	836**	Operating earthmoving equipment; roadbuilding	47
R.	631	Farm or ranch owner	128
	639	Farming: other and miscellaneous	244
S.	661*	Police (public)	151
	666	Fireman	41
	640	U.S. Armed Forces (rank unspecified)	54
	641**	U.S. Armed Forces enlisted personnel	13
Y.	001	(Undecided)	803

*Included in more than one group.

**Tentatively included in the group, on the basis of subjective decision, since the N is too small for conclusive empirical data.

***NEC means "Not Elsewhere Classified."

Significance of the Hierarchical Analysis Results

What significance are we to attach to the fact that there were so many jobs that didn't fall in tight clusters and so few that did? Probably the basic significance of this outcome lies in all that is implied by the apparent nonexistence of a clear hierarchical structure underlying the career groups. The centroids of these career groups are to a great extent scattered widely rather than falling neatly in tight little clusters. The patterns of aptitudes and abilities that characterize various jobs are perhaps almost as diverse as the corresponding patterns for people. Furthermore even within certain occupational groups as identified by job title there are really many quite widely different jobs with widely different patterns of aptitudes and interests required. Within the occupational category M.D., for instance, consider the great difference between what surgeons do and what psychiatrists do. Likewise in the engineer category, civil engineers probably differ more from electronic engineers than they do from those in certain non-engineering occupations. And there are enormous differences, in terms of abilities, interests, and job duties, among those who call themselves computer programers.

Thus not only does hierarchical analysis of occupational groups fail to reveal a unique underlying structure of occupational clusters, but it seems inevitable that any other analytic approach would likewise fail to do so.

An Alternative Hypothesis and an Alternative Approach

It has been implied in the previous section that further efforts to make an analytic procedure reveal the one true set of groups is probably fruitless. But this doesn't mean that we should abandon efforts to group occupations for practical purposes in education and guidance. The practical purposes are too important to be abandoned merely because an obvious and rather widely held initial hypothesis doesn't seem to hold.

If we can't find a unique solution, let us cut the Gordian knot, by recognizing that from a statistical viewpoint there isn't any one best solution but rather a multiplicity of solutions, all about equally good in terms of statistical criteria. And instead of regretting this absence of a definitive solution let us rejoice in it, because it means that we won't be faced with the dilemma of having to choose between a solution which has desirable statistical characteristics but is undesirable from a practical viewpoint and a practical solution for which there is no statistical support. With a vast number of solutions to choose from, all equally sound from a statistical viewpoint, we are free to choose the one that best meets practical demands.

The hypothesis is here proposed that occupational centroids (i.e., sets of means on a number of variables for the members of an occupational group) form a multivariate normal distribution (just as the variables for individuals do). For example if the scores of 15,000 young men from the TALENT sample on 64 cognitive variables constitute a multivariate normal distribution (in 64 dimensions), then the centroids of the approximately 200 occupational categories into which they fall are also likely to constitute a multivariate normal distribution (in the same 64 dimensions). This really seems like a much more plausible and more reasonable assumption than the one that the use of hierarchical analysis implies--namely the assumption that the centroids have a multimodal distribution (with as many modes as there are groups to be formed) rather than a unimodal distribution.

To form groups then, using the proposed unimodal model, all that is necessary is "to chop space up into a number of chunks." All occupations whose centroids fall in the same chunk would be in the same group. The figures which follow show how this would work. All eight figures show the same configuration of points (each point representing the centroid of an occupational category) plotted in two dimensions (because it isn't feasible to plot in 64 or 109 dimensions).

In Figure 1a the plane containing the two dimensions (two variables, F_1 and F_2) has been broken up into squares by a grid. The lines of the grid are parallel to the axes, but there is no reason why they have to be. Figure 1b shows the grid rotated to an oblique angle. Which grid is better would depend on which gives the more useful groups.

Now suppose instead of using squares, we break the plane up into triangular chunks--specifically, uniform equilateral triangles, all of the same size as the squares in Figures 1a and 1b. Figure 2 shows this. Presumably the Figure 2 grouping isn't as good as either Figure 1a or 1b, because triangles are less compact figures than squares, so that one expects the points (centroids) within a triangle to be further apart--in other words to represent less similar

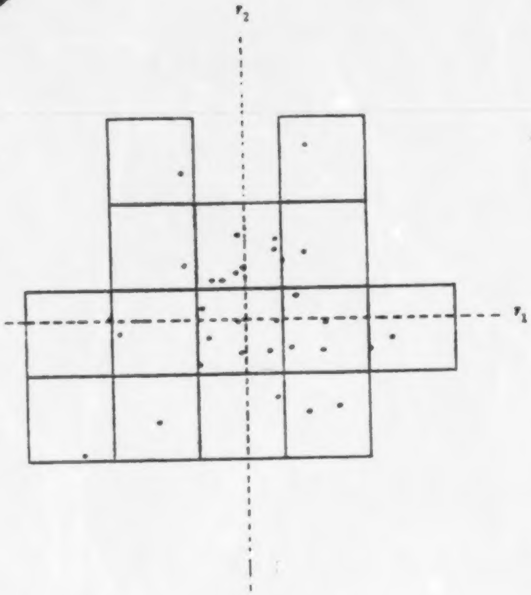


Figure 1a
(14 groups)

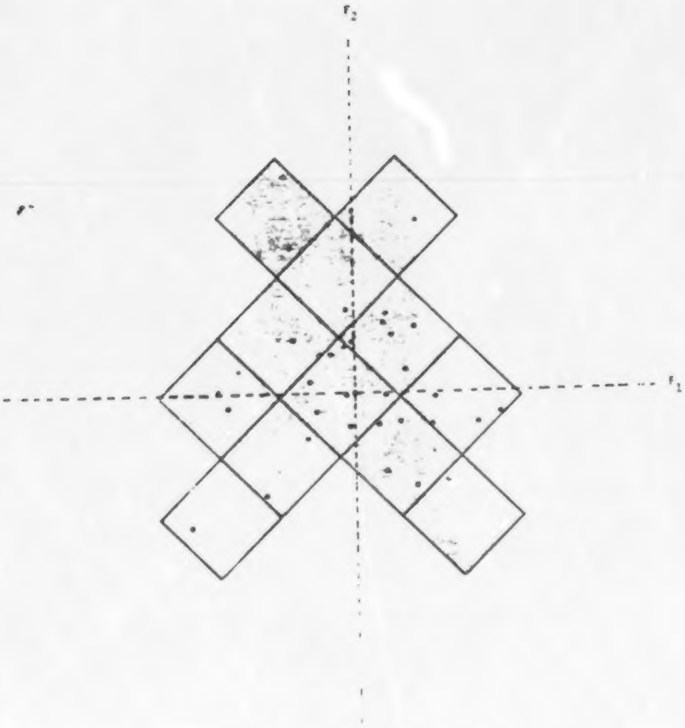


Figure 1b
(12 groups)

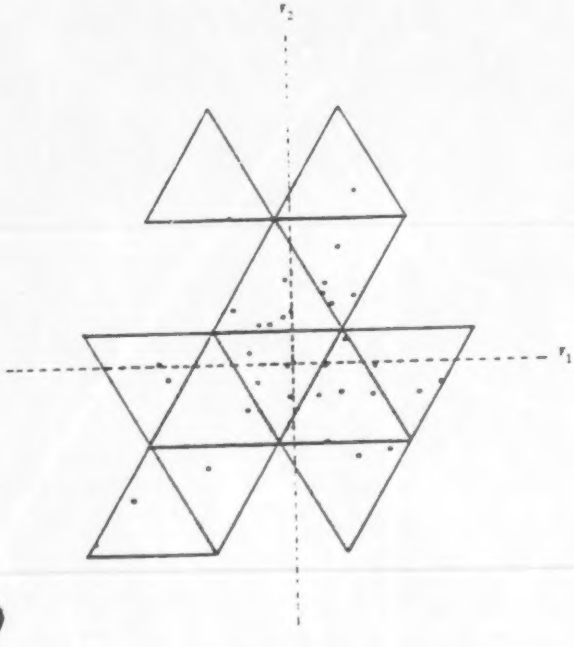


Figure 2
(12 groups)

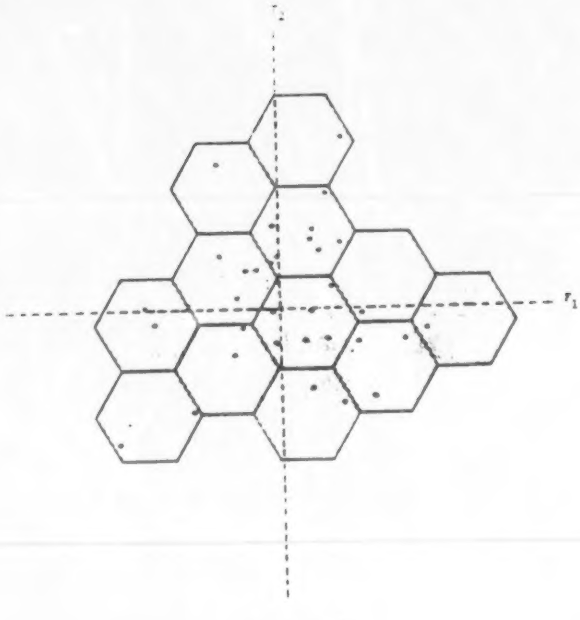


Figure 3
(12 groups)

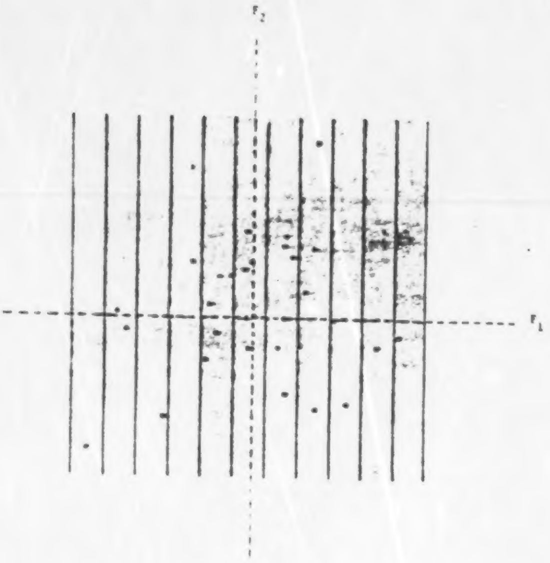


Figure 4a
(11 groups)

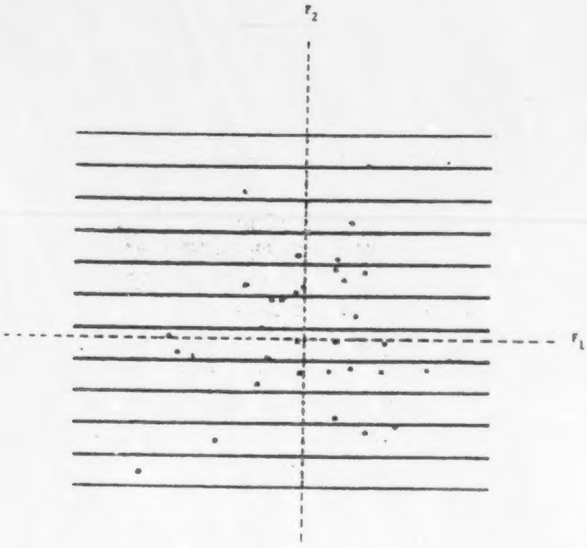


Figure 4b
(11 groups)

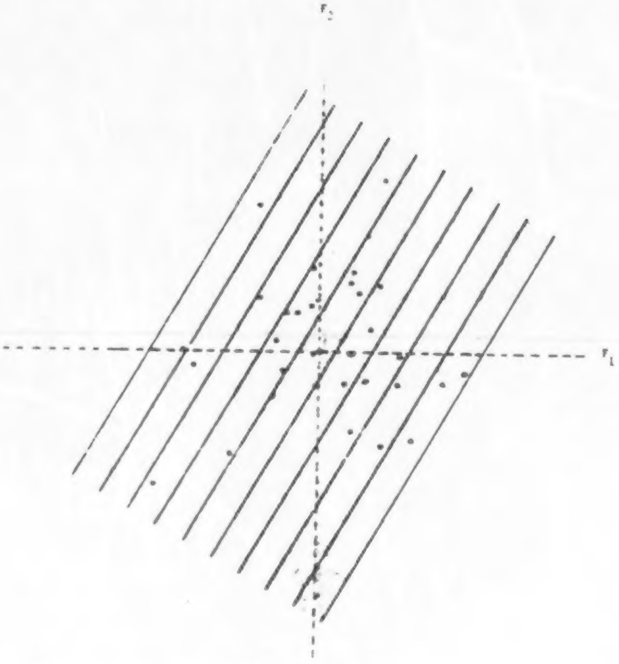


Figure 4c
(9 groups)

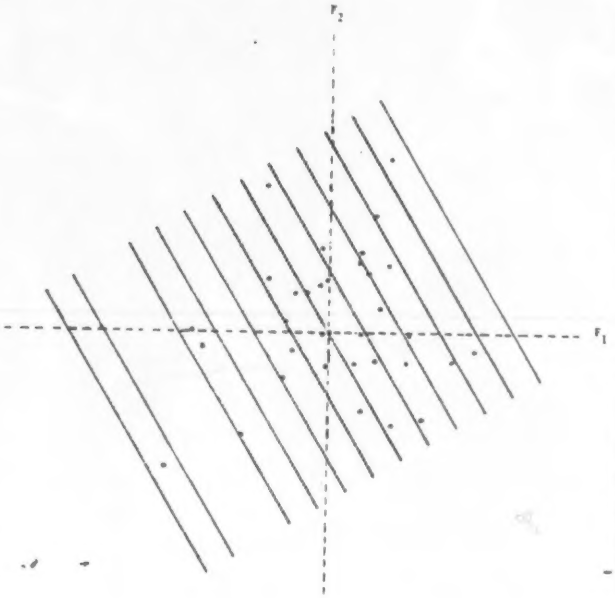


Figure 4d
(9 groups)

groups—than the points within a square of the same size. By the same principle, hexagons should be somewhat preferable to squares, because they are more compact.*

If instead of dividing the space into uniform polygons, such as hexagons, it is divided into narrow strips, as in Figures 4a, 4b, 4c, and 4d, the groups are much less compact, and consequently less satisfactory.

Conclusions

1. In developing ways of grouping occupations, hierarchical analysis is not a promising mode for future exploration. The data are not likely to fit the model.
2. A method has been proposed which explicitly leans much more heavily on practical considerations. It provides a multiplicity of ways of dividing occupations into fairly homogeneous groups, and it provides one basic rule for choosing among these equally sound groupings: choose the way that will be most useful from the point of view of guidance and educational requirements.
3. Occupations don't naturally fall into a very small number of widely separated clusters. There are probably occupations that fit almost any pattern of aptitudes, interests, and other traits that anyone is likely to have. Therefore in selecting a career field there should be less need for a square peg--or even a scalene-triangular peg--to force himself into a round hole than there would otherwise be. Figuratively speaking, the full range of jobs should include lots for scalene triangles of different shapes and size!

---

*Of the regular polygons used in these figures (triangle, square, hexagon) the hexagon is closest to a circle, which is the most compact plane figure there is, geometrically speaking. The regular hexagon is the most compact figure into which a plane can be uniformly divided. There are more compact plane figures--a regular octagon, for instance (to name only one of an infinite multitude)--but a plane cannot be uniformly divided into them.